System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces

Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, Emmanuel Vincent

System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces

Frédéric Bimbot[1], Emmanuel Deruty[2], Gabriel Sargent[3], Emmanuel Vincent[2]

IRISA / Panama Research Group
CNRS, INRIA & Univ. Rennes 1

Author Note

IRISA : Campus Universitaire de Beaulieu, 35042 Rennes cedex, France. Panama Research Group is affiliated to [1]CNRS / UMR 6074, [2]INRIA Rennes Bretagne Atlantique and [3]Université Rennes 1.

Emmanuel Deruty is now with Akoustic Arts and Sony Computer Science Laboratory. Gabriel Sargent is currently post-doctoral researcher at CNAM. Emmanuel Vincent is now at INRIA – Nancy Grand Est.

Electronic correspondence concerning this article should be addressed to frederic.bimbot@irisa.fr, and may be carbon-copied to emmanuel.deruty@gmail.com, gabriel.sargent@yahoo.fr and emmanuel.vincent@inria.fr

Part of this work was achieved in the framework of the QUAERO Research Program (www.quaero.org), funded by OSEO, French state agency for innovation.

This article is based on and extends 2 previously published IRISA Research Reports (RR) [ISSN 2102-6327] : IRISA RR PI-1996 <hal-00713196> (12 pages) and IRISA RR PI-1999 <hal-00868398> (40 pages).

Abstract

This article introduces a model called “System & Contrast” (S&C), which aims at describing the inner organization of structural segments within music pieces in terms of : (i) a *carrier system*, i.e. a sequence of morphological elements forming a multi-dimensional network of self-deducible syntagmatic relationships and (ii) a *contrast*, i.e. a substitutive element, usually the last one, which partly departs from the logic implied by the rest of the system.

With a primary focus on pop music, the S&C model provides a framework to describe internal implication patterns in musical segments by encoding similarities and relations between its constitutive elements so as to minimize the complexity of the resulting description. It is applicable at several timescales and to a wide variety of musical dimensions in a polymorphous way, therefore offering an attractive meta-description of different types of musical contents. It has been used as a central component in the creation of a set of annotations for 380 pop songs (Bimbot, Sargent, Deruty, Guichaoua & Vincent, 2014).

This article formalizes the S&C model, illustrates how it applies to music and establishes its filiation with Narmour’s Implication-Realization model (Narmour 1990, 1992) and Cognitive Rule-Mapping (Narmour, 2000). It introduces the Minimum Description Length scheme as a productive paradigm to support the estimation of S&C descriptions and sketches several tracks where concepts from the domain of Electrical Engineering and Communication Systems can be paralleled with aspects pertaining to the structural description of music patterns by the S&C model.

Strongly based on an Engineering Science viewpoint, the S&C model establishes promising connections between Music Data Processing and Information Retrieval on the one hand, and modern theories in Music Perception and Cognition on the other hand, together with interesting perspectives in other areas in Musicology.

Keywords : music structure, form, implication-realization, cognitive rule-mapping, Kolmogorov complexity, minimum description length, semiotics

System & Contrast : A Polymorphous Model of the Inner Organization of Structural Segments within Music Pieces

Introduction

Context and focus

At low time scales (i.e. typically up to 1 second) music is usually described as a combination of unitary elements such as pitches, durations or chords, drawn from a limited set of pre-defined items. But beyond a certain timescale, music can be furthermore described in terms of piece-specific objects such as cells, motifs, phrases, sections, … whose layout shapes musical segments at increasing scales and ultimately participates to the global *form* of the piece.

Being able to describe the inner organization of musical segments is a very important issue in the context of Music Information Retrieval (MIR), where the automatic processing of music would greatly benefit from an operational model of music structure. Recent studies in MIR have been striving to characterize in different ways structural units and form in music, so as to produce consistent annotated resources for research (Bimbot, Le Blouch, Sargent & Vincent, 2010 ; Bimbot, Deruty, Sargent & Vincent, 2011 ; Smith, Burgoyne, Fujinaga, De Roure & Downie, 2011; Peeters & Deruty, 2009). They have all been facing difficulties in formulating general properties and criteria which could qualify structural units, regardless of the music genre, style or function.

This situation reflects a gap between the profuse literature dedicated to traditional analysis of music structure and form in Musicology (see for instance : Bent & Drabkin, 1998 ; Cadwallader & Gagné, 2011 ; Caplin, 2013 ; Caplin, Hepokowski & Webster, 2009 ; Mac Pherson, 2008 ; Perone, 1998 ; Stein, 1979 ; Zbikowski, 2002) and the need for generic schematic concepts focused on (and suited to) the production of standardized resources usable for music structure analysis in the context of Engineering Sciences (as illustrated for instance in Dannenberg & Goto, 2008 and in Paulus, Müller & Klapuri, 2010). Gradually bridging this gap may not only benefit the development of efficient algorithms for automatic music processing in MIR ; it could also help in defining enhanced concepts for structure analysis in Musicology.

Over the past few years, our research group has been investigating and exploring the issue of music structure description, both from fundamental and experimental viewpoints, with a primary focus on pop music. This work has led to the public release of over 380 annotations of pop songs from three different data sets (Bimbot, Sargent, Deruty, Guichaoua & Vincent, 2014)[1].

The experience acquired through the annotation, discussion and adjudication of several hundreds of pieces has gradually led us to develop a number of concepts and procedures towards versatile representations of music structure, i.e. applicable to a wide variety of music genres (Bimbot, Deruty, Sargent & Vincent, 2012).

In this article, we develop one aspect of this methodology : a description of the *inner organization* of structural segments. By structural segments, we are referring to sections of the musical content at an *intermediate* timescale (typically 10-20 s) constituting potentially relevant units to describe the form of a piece of music at a *large* timescale (namely, its entire span).

[1] These annotations are accessible on `musicdata.gforge.inria.fr/structureAnnotation.html`

Our approach rests upon the observable structure of internal *relations* between the segment's constituents, as the patterns formed by these relations can be assumed to be less specific to a given musical genre and relatively insensitive to the presumed prevalence of particular musical dimensions.

For this purpose, we introduce the System & Contrast (S&C) model. Inspired from algorithmic information theory (Kolmogorov, 1963), a segment is modeled by encoding similarities and relations between its constitutive elements so as to minimize the complexity of the resulting description.

Although the background of this work is that of computational sciences and data processing, we evidence that the S&C model is a straightforward extension of the Implication-Realization model and the Cognitive Rule-Mapping framework, as developed in (Narmour, 2000).

Overview of the article

Section 2 formalizes the S&C model in its standard form (square system), and introduces basic concepts such as those of *carrier system*, *contrast*, *morphological elements* and *syntagmatic relations*. Section 3 develops the links between the model components and actual musical dimensions, properties, elements, relations and patterns in music, and it firmly establishes the affiliation of the S&C model with Narmour's Implication-Realization model (Narmour, 1990 ; Narmour, 1992) and Cognitive Rule-Mapping (Narmour, 2000).

Section 4 and 5 extend the basic (square) S&C model to a more comprehensive set of patterns, able to account for a wide variety of constructions in pop music and beyond. We also discuss how observation-driven structures (such as those described by the S&C model) articulate with cultural conventions (i.e. style structures) as two facets of the music process, corresponding, from a Computer Science viewpoint, to Kolmogorov complexity and Shannon information (respectively).

Section 6 relates the S&C paradigm to a compression scheme which aims at finding the most economical way to describe a musical segment using a MDL (Minimum Description Length) criterion. Section 7 discusses the functions of the S&C model in connection with concepts resorting to Communication Sciences and section 8 points briefly towards the potential of the model in various application domains.

Sixteen examples of musical passages from different musical genres accompany this article to illustrate the S&C model.

Beyond Music Information Retrieval and Engineering Sciences, a number of concepts in this work relate or are inspired from the fields of Cognitive Science, Communication Sciences, Information Theory but also Semiotics and Linguistics. This work should be understood as a contribution towards the formalization of a data model of music patterns, with a multidisciplinary perspective.

The S&C Model : Principles and Formulation

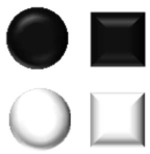
General hypotheses

The main entry for *structure* in the Oxford dictionary states : "*the arrangement of and relations between the parts or elements of something complex*". In line with this definition, the starting point of the System & Contrast model is to consider a musical segment as forming a *system of musical properties.* By *system*, we mean "an interdependent group of items forming a unified whole" (definition of the Merriam-Webster dictionary), but also, "an entity of internal dependencies" (Hjelmslev, 1959).

As a primary hypothesis of the System & Contrast model, we assume the existence of some system of properties within a musical segment, as the essence of the detectable patterns which form its inner organization.

To start, let us illustrate the S&C model with an intuitive presentation, outside the scope of music.

An intuitive presentation



These 4 elements form a *system* based on a combination of two binary *oppositions*, in terms of shape and brightness. We will call this system a *(plain) square system*.

Figure 1 shows a few examples of square systems, for which it is easy to figure out which are the properties used as oppositions, and therefore to *explain* the system. Note that some properties may not participate to the system, as in the 4th example, where the font does not show any systematic behavior.

Figure 1 : Four examples of square systems

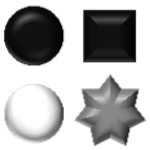
A fundamental property of a square system is its redundancy. Indeed, Figure 2 depicts a few incomplete square systems, where the 4th element is missing, and replaced by a question mark.

Figure 2 : Four incomplete square systems

As can be easily experienced by the reader, some properties of the 4th element are predictable and can be logically *deduced* from the observation of the first 3 elements[1].

As a consequence, it is easy to determine, on the basis of the observation of 3 elements and the presentation of a fourth one, whether this 4th element matches or deviates from the system, and, if this is the case, in what respect.

[1] Respectively : "Germany", a light gray diamond, a very large digit "8" and any item with a NW-SE orientation. In fact, the 4th system in figure 2 could be considered as a complete system : the orientation of the question mark is consistent with that of the rest of the system, and no other property shows any systematic behavior.



These 4 elements now form a *System & Contrast (S&C).* The shape and brightness of the 4th element both contradict the combination of properties expected in 4th position, given the first three elements. The 4th element creates a logical *contrast* within the system.

The characterization of a system and its contrast requires the simultaneous determination of the set of properties which form the system and those which take part in the contrast. Figure 3 illustrates several S&Cs where the properties of the contrast vary over the same baseline *carrier system.*

Carrier system (no contrast)	Brightness-based contrast	Shape-based contrast	Size-based contrast

Figure 3 : Examples of various contrasts, based on the same carrier system

At times termed as "*anticipation by way of induction or analogy*" (CNRS, 1992) (Levy, 2003), the first three elements of the system create a *projective implication* on some of the properties of the fourth element. The contrast appears as a more or less strong *contradiction* of the *expectation* thus arising.

The deviation of contrasting properties on the final element of the system can also be viewed as a *digital modulation* of the *information* conveyed by the carrier system.

Formulation

Using a bi-dimensional indexing notation, a plain square system can be denoted as : $S_0 = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{bmatrix}$

S_0 is assumed to be governed by a network of *similarity* relations :

- horizontal relation : $x_{01} = f(x_{00})$
- vertical relation : $x_{10} = g(x_{00})$
- diagonal relation : $x_{11} = g\big(f(x_{00})\big)$

This can also be stated as a logical proposition :

> x_{11} is to x_{10} what x_{01} is to x_{00}
> **and**[1]
> x_{11} is to x_{01} what x_{10} is to x_{00}

This is nothing else but the *generalization* of the well-known "rule of three", i.e. the relationship between 4 numbers forming a system of proportions.

[1] Provided f and g commute.

Even if many properties are involved in characterizing each element in S_0, the relations f and g may apply only to a subset of these properties, which constitute the *structuring* properties of the system. Reduced to these properties, the carrier system boils down to an initial element (x_{00}) and *two* relations (f, g), the fourth element x_{11} being completely predictable and therefore redundant.

Following similar notations as above, the "System & Contrast" can be noted : S = $\begin{bmatrix} x_{00} & x_{01} \\ x_{10} & \bar{x}_{11} \end{bmatrix}$

Matrix S now requires a specific diagonal relation to describe the contrast : $\bar{x}_{11} = k(x_{00})$, with $k \neq gof$.

The contrast results from the *disparity* between k and gof, which can itself be viewed as a relation γ which expresses the deviation of the actual element $\bar{x}_{11}$ from the (virtual) *expected* one $x_{11} = (gof)(x_{00})$. In mathematical language : $\gamma = ko(gof)^{-1}$

We now have the following situation :

- $\bar{x}_{11}$ is **<u>not</u>** to x_{10} what x_{01} is to x_{00}
- **and/or**
- $\bar{x}_{11}$ is **<u>not</u>** to x_{01} what x_{10} is to x_{00}

Element $\bar{x}_{11}$ is breaking the anticipation triggered by x_{00}, x_{01} and x_{10} and creates a logical disruption. Relation γ thus appears as a discordance in the system, which can be detected in reference to the other elements by first <u>deducing</u> and then <u>factoring out</u> the structuring properties of the underlying *carrier* system S_0.

Analyzing a S&C

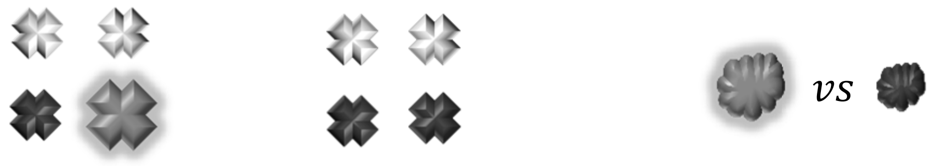
The structuring properties involved in the description of a S&C can rest on virtually <u>any property</u> or combination of properties, provided they evolve in an *organized* manner and form a consistent set of detectable relationships.

Consider the following two quadruplets :

S&C #1 S&C #2

Figure 4 : Examples of two S&Cs (analyzed in the text).

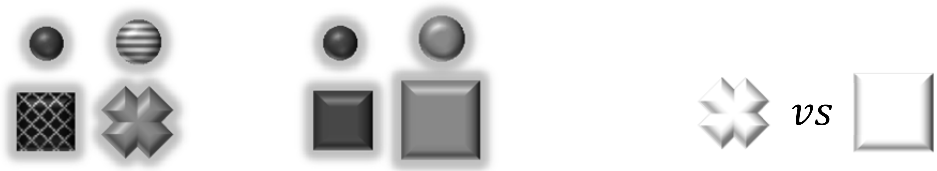
In S&C #1, the 5 properties needed to describe all the observed elements are *shape, size, brightness, halo* and a subtle *shade orientation*. Given the first three items, the logical element in position 4 would be a *dark gray cross* of the *same size* as the others, illuminated from "*NW*", with *no halo*. It is indeed a *cross* illuminated from *NW* but it is *large(r), medium gray,* and surrounded *with a halo*. As summarized on Figure 5, the contrast affects 3 properties : *size, brightness* and *halo*.



System & Contrast | Carrier system | Contrast (reduced form)

Figure 5 : Result of the analysis of S&C #1

In S&C #2, *shape, brightness, size, halo, texture* and potentially *orientation* are properties of the elements in the system. However, (i) *texture* varies erratically and can therefore be considered as an off-system property and (ii) the status of *orientation* as a systematic property is not decidable, as it is not possible to evaluate it for the circles. Among the remaining properties, only *shape* and *size* participate to the contrast : the 4th element is a *large cross* instead of being a *very-large square*. Figure 6 summarizes the result of the analysis of S&C #2.



System & Contrast | Carrier system | Contrast (reduced form)

Figure 6 : Result of the analysis of S&C #2

Redundancy in the system offers the possibility to "calibrate" the contrastive information in the 4th position, which can be deduced from the way the properties of individual items locally <u>vary across the system</u> and create patterns (or not). As a consequence, a same element in 4th position can have very different *contrastive values* in 2 distinct systems.

S&C description

Figure 7 below depicts a square S&C, unfolded over time, on which the main constituents of the system are represented : the morphological elements x_{ij}, the syntagmatic relations *f*, *g* and the contrast function γ.

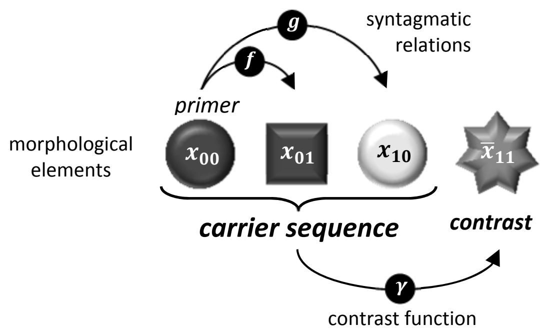


Figure 7 : A schematic view of a 2×2 square S&C model components (unfolded form) on a non-musical example

The prototypical systems considered so far are 2×2 (square) systems of 4 *morphological elements*, but further on in this article, we will also consider systems with other *configurations*. The first element x_{00} is called the *primer* and the last one, the *contrast*. If γ is the identity function, there is no contrast at the end of the system, and the S&C is said to be *plain*.

S*yntagmatic relations f* and *g* account for correspondences between elements within the system. They can be understood as exact or approximate *similarities,* or as latent *transformation rules* which map one element to another. They are multi-dimensional, i.e. they can apply to several variables or properties at a time, but for independent variables, they can be de-composed on each dimension separately.

We call the quadruplet (x_{00}, f, g, γ) the S&C *description* of S, which can be viewed metaphorically as the "genetic program" of the system[1].

Note that the S&C model enforces a *causality principle* in the sense that the *direction* of the relationships between elements within the system is assumed to be in accordance with the order in which these elements occur in the unfolded system. As a consequence, the contrast is always in final position, although, in some cases, it would also be possible to explain the system on the basis of a disparity affecting another element than the last one.

Modelling Music Segments with the S&C

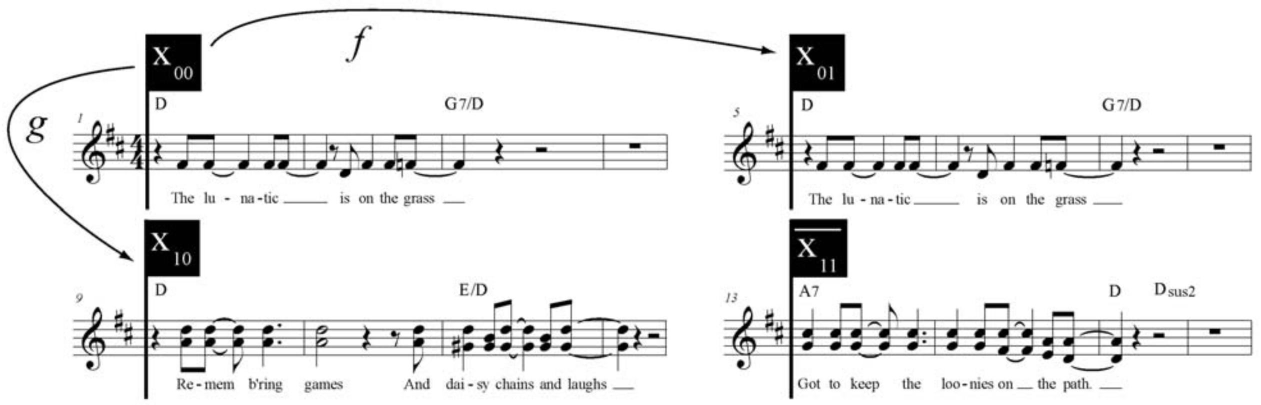
Music appears mainly as a *sequential* and *dynamic* presentation of acoustic (or symbolic) material. However, the S&C model assumes *matricial* relationships between *discrete* objects. Analyzing a musical content as a S&C therefore involves implicit operations of *delinearization* of the musical flow and *discretization* of its properties.

In this section, we illustrate a few instances of S&Cs on actual musical segments and discuss the role of the S&C model in their description. As morphological elements (namely the x_{ij}), we consider musical motifs of a few seconds (typically 2 bars, but occasionally larger or smaller fragments).

Some examples

Example n°1

Pink Floyd – Brain Damage (Composer : Roger Waters)
The Dark Side of the Moon, EMI 1973. Timing : 0'15-0'43
"Pink Floyd : The Dark Side of the Moon, Guitar Tablature Edition" pp. 109-111. Published by Music Sales America, 1992



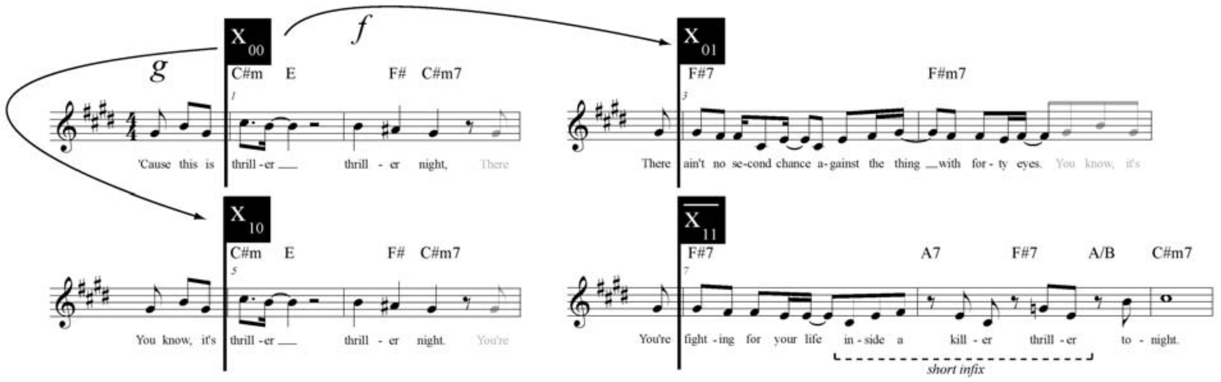
From a pop song from the 70's, Example 1 appears as an $[aabc]$ sequence, a very common pattern in music, which we call *sentence-like*. While relation f is the exact identity function, relation *g* consists of a major reorganization of the second part of the motifs X_{00} and X_{01} : the addition of a second vocal line, and denser melodic profiles repeated with an internal transpo-

[1] Echoing Narmour's "*genetic code*" of melody (Narmour, 1989).

sition. The last segment $\overline{\mathrm{X}}_{11}$ acts as a contrast as it departs from the repetition of X_{10} : it appears as a less drastic transformation of the initial motives, sharing some properties of X_{10} (the addition of a second vocal line) but with a melodic placement which almost falls back to that of the primary element X_{00}. Therefore, the whole segment can be viewed as a S&C resulting from the congruence of (at least) two patterns : $[aaa'c]$ (harmony) and $[aaba']$ (rhythm), both participating to the observed form[1].

Example n°2

Michael Jackson – Thriller (Composer : Rod Temperton)
Thriller, EMI 1982. Timing : 2'26-2'40
"Thriller", pp. 25-26, Published by Rodsongs (PRS), 1982



From a famous pop song from the 80's, the 9 bar segment of Example 2 illustrates an interesting case of an $[abac]$ (*period-like*) pattern. Relation f can be seen as introducing *new* musical material in X_{01}, but X_{10} presents again the material of X_{00} (i.e. $g = id$ except for the beginning of the lyrics). The contrast γ can be viewed as composite and multidimensional. On bar 7 in $\overline{\mathrm{X}}_{11}$, γ is almost *identity*, as this sub-segment starts very much like X_{01}. Then come 2 bars which show a clear disparity with the course of X_{01}, with the introduction of a new, heavily syncopated motive followed by a completely steady note. Moreover, $\overline{\mathrm{X}}_{11}$ develops over 3 bars, which also creates a contrast of duration with the other morphological elements (2 bars). A component of the contrast γ therefore consists in a stretching effect, by the insertion of musical matter at the level of bar 8 (marked as "*short infix*"), which delays the conclusion of the sequence (some sort of *phase shift* in the resolution of the segment).

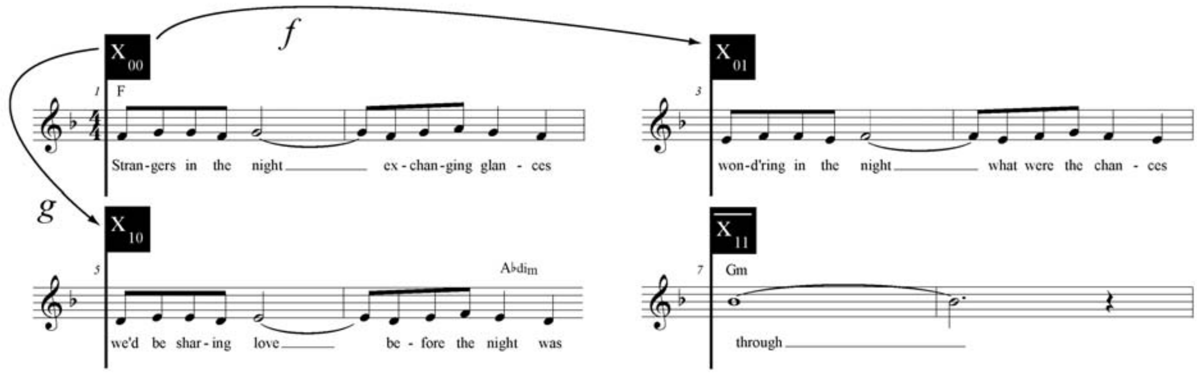
In Example 3, the first verse of this well-known love song of the 60's illustrates a S&C type which does not correspond either to a sentence-like, nor to a period-like pattern. X_{01} and X_{10} are best described as two successive diatonic transpositions of the primary element X_{00}, while the harmony remains constant. Here, we have $g = f^2$, but the last 2 bars $\overline{\mathrm{X}}_{11}$ introduce a completely distinct element, made of a single note that lasts seven beats and whose pitch is significantly higher than all previously heard pitches. $\overline{\mathrm{X}}_{11}$ appears as a complete contrast with the progression installed by the three previous elements, in terms of rhythmic pattern, melodic pitch and shape, chord root and mode, musical and vocal information flow, … It can be denoted as an $[a^0a^1a^2b]$ pattern, i.e. a *broken progression*[2].

[1] Here the morphological elements considered have 4 bars. In this example, sub-systems also exist at the 2-bar scale, where bars 1-8 form as a plain system $[abab]$ followed by a $[a_2a_3a_4c]$ pattern over bars *9-16.* At an even lower time-scale (1 bar), a good approximation of each 4-bar group could be $[abcc][abcc][a_1b_1a_2b_2][a_3b_3cc]$. This is a first illustration of how the S&C model can operate jointly at several time-scales.

[2] Let us point out that, whereas f and g introduce little novelty in this example, γ proposes something completely new. This is to be put in perspective with the two previous examples, where f or g introduced much more novelty, while γ tended to be less innovative.

Example n°3

Frank Sinatra – Strangers in the Night (Composers : Kaempfert, Singleton, Snyder)
Strangers in the Night, Reprise, 1966. Timing : 0’11-0’32
“Strangers in the Night”, p. 2, Universal Music Publ. Group / Hal Leonard Corp., 1966-2011



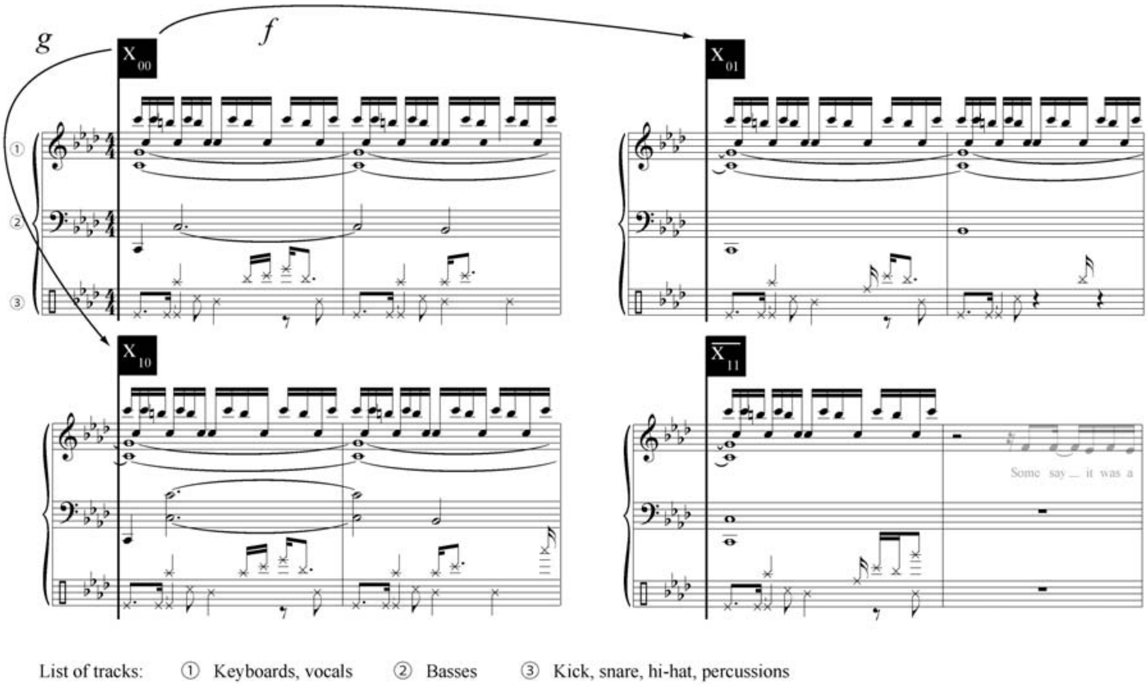
In music, melody, harmony and rhythm frequently play a predominant role in conveying or signaling structural information. But considering that potentially <u>any</u> musical dimension can contribute to the inner structure of a musical segment makes it possible to approach a wider range of musical contents within the S&C framework and in particular, pieces or passages for which the “usual” musical dimensions do not prevail in their inner organization. This is illustrated in the two examples hereafter.

Whereas, in many cases, the contrast function is complex and applies diversely over various musical dimensions, Example 4 is an instance of “industrial music” where a more radical approach can be observed : the contrast consists in a total *suppression* of the content of the second part of the 4th morphological element $\overline{\mathrm{X}}_{11}$, resulting in a sudden and complete silence on all musical dimensions. This creates a definite effect of surprise, leaves room for and focuses attention on the deployment of the anacrusis of the forthcoming segment. Altogether, the segment tends to follow a period-like behavior $[abac]$, with γ being *id* over the first half of the last element and *zero* over its second half ($c \approx b'$).

Taking into consideration, in Example 5, only the pitched instruments (basses) and the traditional drum section (kick, snare, hi-hat), this segment of electronic music appears as a sequence of four identical elements, namely $[aaaa]$. However, careful listening reveals the presence of a set of light percussive samples organized into a period-like system $[abac]$. Function f associates the pattern heard in the *primer* X_{00} to a sequence of four syncopated, regularly spaced hits, along with another syncopated hit near the end of bar 3. The function g is the identity, and the contrast γ introduces yet another completely new motif in track 1. In this particular case, the most salient instruments exhibit a completely monotonic pattern, while the contrastive component in the S&C develops over non-conventional musical elements in an almost inconspicuous way.

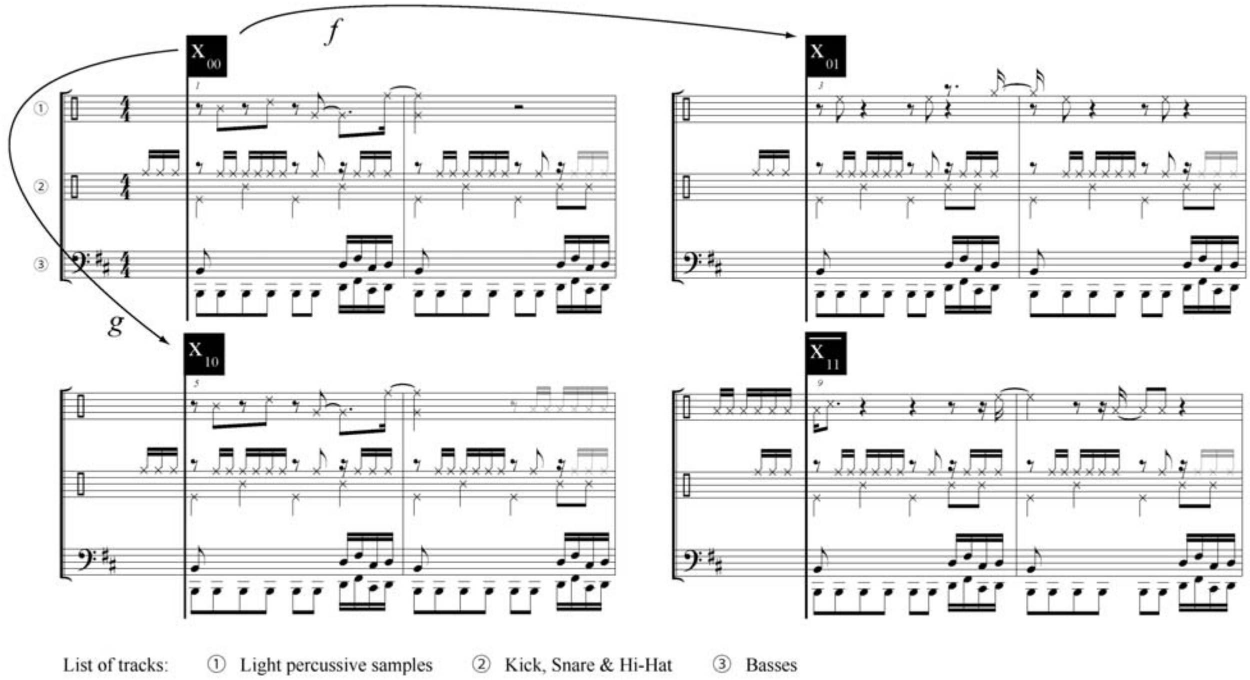
Example n°4

Nine Inch Nails – The Warning (Real World Remix)
(Composers : Trent Reznor, Stefan Goodchild, Doudou N'Diaye Rose)
Y34RZ3R0R3M1X3D, Interscope, 2007. Timing : 0'27-0'46. *Transcribed by ear*



Example n°5

Olivier Lieb – Epsilon Eridani
Epsilon Eridani EP, Bedrock Records, 2011
Timing 1'17 – 1'32. *Transcribed by ear*



Note that, in the five examples above, the S&C descriptions are presented in a static manner. However, listening to music is inherently a dynamic process, radically different from looking at a matrix of images all at once. The S&C approach thus assumes that musical properties and their relations are encoded and memorized on-line, possibly giving rise to multiple competing hypothesis. Then, once the entire passage is complete, the listener is able to figure out which musical parameters are actually relevant to the S&C or not (and this does not in any way dismiss their relative relevance w.r.t. other musical aspects).

S&C as a meta-prototype

The S&C model is designed and formulated so as to be applicable in a versatile way to multiple musical dimensions and to a wide variety of music pieces. As a neutral level of analysis (Nattiez, 1987), the proposed approach does not seek to decipher and uncover the *message* behind a musical segment. It only aims at providing a standardized description of organizational patterns in structural segments[1].

The S&C framework approaches the description of structural segments as a *model matching* problem, where morphological elements, structuring properties and syntagmatic functions are <u>jointly estimated</u> so as to optimize the explanation of the observed data within the class of S&C models.

This conception can be related to prototype theory addressed in (Deliège, 2001) and particularly to the notion of "abstracted central tendency", as defined in (Lamont, 2001) : *"the prototype is viewed either as a particular privileged exemplar of a given category or as an abstracted central tendency, and similarity is a function of the distance between a given item and the prototype, measured in terms of common and distinctive features.*

In the S&C approach, however, the adequacy of the data to the prototype is not expressed as a similarity measure but as an empirical cost that rewards the goodness of fit of the observations to the particular S&C model that can be inferred from the observations themselves. In this sense, rather than a set of exemplar or average reference patterns, the S&C appears as a template at a higher level of abstraction, some sort of meta-prototype used to schematize the musical content and to gauge the quantity of information that can thus be explained from the data.

Encoding S&C information

Structuring properties in a S&C emerge as a consequence of their *relative behavior* within the system : identifying them is entirely part of the S&C analysis process. This view happens to be particularly productive for pop music, and especially urban music and electronic music, for which conventional musical dimensions (and in particular, harmony) may be of little help to explain and characterize structure.

For music, a non-exhaustive list of possible structuring properties is :

- Melody contour / melodic intervals / support notes of the melody / sign of variations of the melody,…
- Underlying harmony / chord sequences / root progressions,
- Rhythmic placement / rhythmic cells and patterns,
- Pauses / energy distribution and flow,
- Drum sequences and loops, …
- Rhymes / phonetic flow / chant, …
- Instrumental timbres / arrangements and support / special effect schemes, …

[1] In Ruwet's conception (Ruwet, 1966), this can be viewed as a semiotic analysis process, focusing on the *structure* of the *code* and its discovery.

- Macroscopic properties, such as mode, tonality, tempo[1],…
- Sub-structural properties (such as the size) of motivic elements,
- etc…

These properties may not vary independently, and several of them co-vary in a systematic way.

Typical syntagmatic relationships can be functions which operate on the harmony, melody, percussions, note placement…

- identity (exact or almost exact repetition),
- chromatic transposition (constant shift in half-tones on the chromatic scale),
- diatonic transposition (constant shift in degrees on the current scale),
- mode or scale change (same degree(s) on a different scale),
- changes in note duration and placement,
- beat pattern alteration / inversion / complementation,
- etc…

… but they may also apply to the amplitude, time or timbral dimensions such as :

- amplitude increase / decrease / silencing (i.e. zeroing the amplitude)
- rhythmic scale change / phase shift
- fragmentation / augmentation
- extension / simplification (i.e. insertion / deletion of auxiliary musical material, such as ornaments)
- adjunction / suppression / change of instrument(s)
- etc…

We summarize in Table 1 different archetypes of relations between elements in the system, together with their abstract codification.

Type of relation	**Notation**	
	First element	Second element
Both elements are identical (exact repetition)	a	a
Both elements are not considered as significantly different, or their differences are not deemed relevant to explain the system	a	$\tilde{a}$
The second element is obtained by shifting the first one along some parametric scale (for example, a transposition)	a or a^0	$a^1, a^2, a^3, \dots$
The two elements are dual, i.e. they are reciprocal images by some sort of symmetry, inversion or complementation (note that $a^{**} = a$)	a	a^*
The second element is a strengthened/lengthened (resp. weakened/shortened) version of the first one	a	a^+(resp. a^-)
The two elements start (resp. end) alike but they end (resp. start) differently (note that aa'forms a S&C at the immediately lower scale $a_0a_1a_0\bar{a}_1$)	a	a'(resp. a'')
The two elements are related by a specific similarity relation or mapping transformation, not covered above.	a	$f(a), g(a), \dots$
The two elements as considered as distinct, or no particular relation can be invoked to relate them within the system	a	$b, c, \dots$

Table 1 : Abstract codification of similarities between musical elements

[1] Such properties usually vary at a slow pace but they can become “structuring properties”, if they create patterns at the working scale.

The contrast is also modeled as a function γ, which operates by convention in reference to the combination *gof*. Function γ can be subject to considerable variability and the intra-opus repetition of a same carrier system with different contrasts is an extremely frequent narrative strategy, within and across many music genres. Occasionally, $\gamma = id$, which leads to the plain realization of the carrier system (no denial of the expectation on any musical dimension), i.e. a S&C with no contrast.

In many cases, morphological elements are of equal size[1], but situations arise where elements are of different sizes (such as in Example 2). In fact, the size of the morphological elements can be one of the properties which governs the system and/or its contrast. S&C descriptions based on regularity are *a priori* preferred but variations in element size can be one of the properties that creates a pattern.

Implication-Realization, Cognitive Rule-Mapping and the S&C model

As it now stands out from the above presentation, the S&C model relies on three major hypotheses :

(i) congruent similarities between elements in a musical segment create a network of relations which contributes to the perception of its structural cohesion,

(ii) by creating some redundancy, these relations develop a system of projective implications, usually over multiple musical dimensions,

(iii) the realization of the contrast constitutes a more or less strong denial of the built implications, which flags up and participates to the closure of the segment, by concluding an expectation process.

Originally designed for the analysis and cognition of melody structure and melodic complexity (Narmour 1990 ; Narmour 1992), the Implication-Realization (I-R) model was later extended by its author (Narmour 2000) towards the concept of Cognitive Rule-Mapping (CRM). In this section, we show that the S&C model rests on hypotheses compatible with those of the I-R model and we formally establish the S&C as a straightforward generalization of the CRM scheme.

In its initial version, Narmour's I-R model (Narmour, 1989) is based on two fundamental principles, summarized by the two famous formal hypotheses :

$$A + A \xrightarrow{implies} A$$
$$A + B \xrightarrow{implies} C$$

In words, these 2 rules state that two similar items trigger the expectation of a repetition of that same item, whereas two distinct items prompt the implication of a subsequent change in the next item. Narmour defines closure as the *"termination, blunting, inhibition or weakening of melodic implication brought about by its realization or its denial"* (Cross, 1994) and relates a number of archetypical melodic processes to different degrees of closure. Note that, in this form, the I-R model is a second-order model, in the sense that it defines projective implication as a function of the previous two elements.

Narmour (2000) introduces a more general set of concepts under the title "*Music expectation by cognitive rule-mapping*".

[1] We term the ratio between the size of the segment and the (typical) size of the morphological elements the *granularity* of the S&C. Another term could have been *resolution*, but it would have been a source of ambiguity given the usual meaning of that word in music.

The author develops and illustrates the interactions between (i) rule inference arising from similarities across successive musical elements and (ii) how the inferred rules support subsequent implications. This leads to a more sophisticated set of implications, invoked and explained by Narmour throughout his article and listed below in Table 2 (using Narmour's own formalism[1]).

$\underset{(1st\ form)}{A^0} + \underset{(2nd\ form)}{sA^0} \xrightarrow{implies} \underset{(2nd\ form)}{A^0}$	[Rule #1] (p. 334)
$\underset{(1st\ form)}{A^0} + \underset{(2nd\ form)}{A^0} \xrightarrow{implies} \underset{(3rd\ form)}{A^0}$	[Rule #2] (p. 334)
$A^0 + A^1 \xrightarrow{implies} A^2$	[Rule #3] (p. 336)
$A^0 + T_n(sA^0) \xrightarrow{implies} S_n(A^0) = A^1$	[Rule #4] (p. 337)
$A^0 + T_n(s_2A^0) \xrightarrow{implies} S_n(s_2A^0) = A^1$	[Rule #5] (p. 343)

Table 2 : Set of implication rules from (Narmour, 2000).
sA^0 (and s_2A^0) denote subsets of form (or properties of) A^0,
T_n transpositional rules, and S_n the corresponding sequential output.

Rule #2 is nothing more than the basic repetition rule of the original I-R model : "$A + A \rightarrow A$".

Rule #3 is a generalization of rule #2, corresponding to situations when "*the listener may cognitively understand, albeit unconsciously that […] the varied part of a given repetition may invoke an organizing rule*" (p. 335). Iteration of the rule is then projected onto the next element, and Narmour points out the feedback between the two : *"repetition constrains the rule, but rule also constrains the repetition (see Jones, 1990)*". However, rule #2 and #3 remain second-order implication rules, in the sense that they inform on projective implications based only on the previous 2 elements.

Rule #1 deals with the implication that arises on a second form, when a "subset" of that second form is identical to its counterpart in the first form. Narmour explains rule #1 in the following terms (p. 334) : "*A subset (s) from the second form exactly mimics the first A^0 and thus triggers the expectation of an exact repeat (the second form)*".

Decomposing A^0 into two sub-segments $A^0 = ab$, and considering a as a subset of A^0 (i.e. $sA^0 = a$), the application of rule #1 yields : $ab + a \rightarrow b$, or, more conveniently : $a + b + a \rightarrow b$. We thus obtain a third-order implication rule, which says that if a first occurrence of a has been followed by b, a second a triggers the expectation of a second b. It is worth noting that this implication is not in contradiction with $a + b \nrightarrow a$ being a denial : the third-order relation $a + b + a \rightarrow b$ simply means that once the sequence aba has actually been observed, the projective implication on the next element becomes b.

Let us now consider rule #4, which is fundamental to our demonstration. The steps involved by Narmour in the formalization of this rule are literally (p. 337) :

[1] … thus consolidating Narmour's hope "*that the mathematical formulations of the musical rules will be useful to those interested in computational modeling"* (Narmour, 2000, p. 331).

> *(1) perceive that the initial subset of the second form (sA^1) is similar and fundamentally analogous to the beginning of the first one (sA^0)*
> *(2) abstractly code the variables of A^1 so as to access the relevant rule, and*
> *(3) then apply the rule to project the sequential continuation*
> *(Narmour, 2000)*

Denoting again $A^0 = ab$ and $A^1 = a_1 b_1$, step 1 consists in identifying the similarities between a and a_1, step 2 describes the inference of a mapping rule between the 2 elements (say, a function *g*) and step 3 applies the inferred function to the rest of A^0 to project the implication, namely :

$$a + b + g(a) \rightarrow g(b)$$

This rule (which is also a third-order rule) tells us that the implied (4th) element is to the third one as the second one is to the first one. Assuming now another mapping rule $b = f(a)$, we get a general implication scheme, which exactly corresponds to the S&C model in its plain square form :

$$\boxed{a + f(a) + g(a) \xrightarrow{implies} g(f(a))}$$

Ultimately, rule #5 is introduced by Narmour to account for "*bifurcated streams*", which occur when a "*sequence involves rule iteration applied to segments alternating with mappings of registral return*", i.e. $ab^0ab^1ab^2\dots$ (denoted as $A^0A^1A^2$ at the immediately upper scale). The subset of properties s_2 subjected to implications are the successive pitch intervals $[b^0 - a], [b^1 - a], [b^2 - a]$ and rule #5 can be viewed as the application of iterative rule #3 ($A^0 + A^1 \rightarrow A^2$) specifically to these structuring properties[1,2].

Let us consider Example 6, drawn from Narmour's article, and commented by the author as follows : "*in measure 3, Franck sequences the opening skip of the fourth (F-Bb) to a skip of a sixth (F-Db), but then with the return of the F-Bb he denies that expectation (that the leap will expand triadically to an octave)".*

Narmour focuses his analysis of this passage on a denial relation which can be written as : $A^0 + A^1 \nrightarrow A^{0'}$, where $A^{0'}$ represents segment "$A^0 \dots B$" and the parametric superscript, the interval of the "skip". Narmour's analysis stops here. However, while B surely creates some partial denial (say, at a 1-bar scale), the implication process does not halt there : a second, stronger denial clearly arises from C over its entire duration (~ 2 bars)[3] : $A^0 + A^1 + A^{0'} \nrightarrow C$ (third-order implication, denied). Moreover, notation $A^0 \dots B$ suggests that the third segment relates more directly to A^0 rather than to A^1, in line with an implicit matricial organization of the musical content.

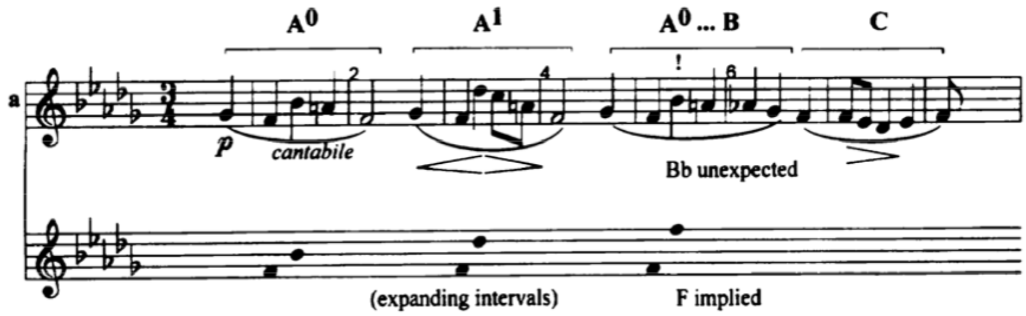
---

[1] Strictly speaking, rule #5, as we understand it, should write $A^0 + [T_n(s_2A^0) = A^1] \rightarrow S_n(s_2A^1) = A^2$, to make it clear that the implication concerns a *third* form, A^2.

[2] As a support to the proposed rule, Narmour provides examples consisting in alternating patterns of intervallic increments and identifies the main implication process as built on the successive transpositions on the *second part* of the form (s_2A^0). This situation can also be accounted for by an extension of the S&C model to a system of 3x2 elements : $[a + b] + [f(a) + g(b)] \rightarrow [f^2(a) + g^2(b)]$ (see further, section on hexadic systems).

[3] In particular, the melody contrastively returns to F with an ascending movement.

Example n°6 **César Franck –** Symphony in D Minor, II. 17-24
Reproduced from Narmour (2000), p. 353



Together with the principles of Cognitive Rule-Mapping, Narmour (2000) points towards its extension to a variety of musical dimensions and mapping functions : *“Other parameters appear amenable to cognitive rule-mapping as well”* (p. 331). He thoroughly illustrates this by examples (pp. 365-372), where *tempo*, *pace*, *texture*, *harmony, dynamics* are identifiable as “*other parametrical style shapes* [to which] *iterative but largely unconscious rule-mapping can cognitively apply*” (p. 366). He shows how these musical dimensions can form patterns such as “*systematic increase in pace*” (p. 368), “*dissonance : increase … decrease … increase … decrease*” (p. 369), or “*harmonic syntax contrasted with rule-mapped dissonance*” (p. 372). Here again, similarity detection and rule-mapping on musical elements are invoked and combined to build projective implications. Although Narmour provides examples from art music, these principles appear to be particularly wide-spread in pop music.

All of above developments firmly root the S&C model in the direct lineage of Narmour’s I-R model and CRM principles. Narmour’s work establishes clearly the role played by projective implications and their denial, in the organization of musical matter. We evidence here that the S&C model extends explicitly the rules of the I-R model to higher order dependencies and formally generalizes its principles to multiple musical dimensions and various classes of mapping functions, in line with the CRM principles.

The S&C model also enables the formalization of the cognitive rule-mapping process in a matrix framework thus providing a multi-dimensional view of the relational processes that develop within a musical segment. This permits to natively account for alternating similarities across properties and elements which are not necessarily contiguous.

Hence, the S&C model can be viewed as a framework which translates, extends and encapsulates the I-R model and the CRM scheme into a formal and rather generic data structure.

S&C Morpho-Syntagmatic[1] Patterns

In the S&C model, the range of musical dimensions and syntagmatic functions which can be invoked to describe the system and its contrast is potentially very large, and their relative prevalence and dependencies are surely complex to model.

However, the picture gets simpler when focusing on functions f, g (and γ) limited to a binary set *{id, new}*, indicating re-

[1] The term *morpho-syntagmatic* is chosen in analogy with structural linguistics (Chomsky, 1957), where it refers to a purely grammatical approach of the message based on the form and relations of its constitutive elements, outside of any functional (or semantic) considerations.

spectively whether an element in the system is judged *similar/analogous* or *distinct/unrelated* to the primer a (and, for the contrast, whether it is similar or distinct to its non-contrastive implied form). This temporarily restricts the analysis of S&Cs to the most salient structuring dimensions, but this turns out to be effective in many situations, in particular for pop music.

Square systems

Based on the third-order implication model, the list of 2×2 patterns in Table 3 can be readily derived :

Plain	**Contrastive**	
$aaaa$	$aaac$	
$aabb$	$aabc$	$aaba$
$abab$	$abac$	$abaa$

Table 3 : Primary repetition-based patterns encompassed by the S&C model

These patterns correspond to common configurations in music, at different timescales ranging between a few seconds (sometimes even less) up to 25-30 seconds (or occasionally more). They can be considered as unambiguous S&C patterns as it is unequivocally possible to determine their underlying carrier system (column "Plain"). They can easily be combined with transformations from Table 1, to refine their description : $[a\tilde{a}bc]$, $[aba'c]$, $[a^0a^1bc]$, $[abab^*]$, $[aaa^+a^-]$, etc… Note that patterns $[aaba]$ and $[abaa]$ are particular cases of contrastive forms, where the last a is denying an expected b (in the plain form).

The "strength" of a S&C results from the <u>synchronous</u> (or congruent) realization of <u>several</u> such primary patterns over distinct musical dimensions (while some other musical dimensions may not follow any pattern[1]).

Six other patterns can be obtained as the combination of 2 or 3 distinct elements (see Table 4). They may also be described as S&Cs, but they are potentially ambiguous : it is not always possible to tell whether they are plain or contrastive. For instance, $[abbc]$ can be judged as *non*-contrastive only if the function assumed to relate b to a also relates c to b <u>in the same way</u>. This may not always be easy to arbitrate and these patterns are considered as weaker in the context of the S&C model, because they tend to support a less well-defined structural organization.

Contrastive	**Ambiguous**
$abbb$	$abbc, abba$
$abcb, abcc$	$abca$

Table 4 : Secondary repetition-based patterns encompassed by the S&C model

In the forthcoming subsections, we briefly develop how the S&C model can be extended to other configurations, beyond the prototypical 2×2 phrasal structure.

Dyads and triadic systems

Dyads (i.e. sequences of 2 elements) appear as repetitions aa, semi-repetitions aa' or oppositions ab. They may be considered as a 1-dimensional "system" at that scale (underlain by a first-order projection rule $a \rightarrow a$), but they can generally

[1] For instance, in a structural segment, the chords may go $[abac]$, the drums $[aaab]$ and the lyrics $[abab]$, while the melody goes $[abcd]$.

be described as a two-dimensional (square) S&C at the immediately lower scale. Conversely, the repetition of a dyadic system $[abab]$ forms a non-contrastive square system at the immediately upper scale.

Triadic systems (i.e. systems of 3 elements) are central to Narmour's I-R theory ($a + a \to a$) and they can be viewed as the simplest form of systems. Based on the repetition of a single mapping function *f*, they yield an iterative carrier form : $a + f(a) \to f^2(a)$. When f is binary, the typical triadic carrier system writes as $[aaa]$ in front of which $[aab]$ appears as a contrastive form (when $b \neq a$). However, triadic patterns such as $[aba]$ and $[abb]$ are also contrastive, as in both cases the last element acts as a denial.

Some triadic segments can also be analyzed as particular cases of *square* systems, where one element is missing. Out of context, it may be difficult to decide whether a sequence of 3 elements is a "true" triadic system or a truncated square system. However, the latter hypothesis can be privileged when there exists, somewhere else in the piece, the realization of the whole square system, to which the truncated system directly relates.

Pentadic sequences

Although not as frequent as square systems, pentadic configurations (i.e. structural segments formed of 5 morphological elements) are rather common in music. Interestingly, apart for the trivial sequence $[aaaaa]$, any other sequence of 5 elements containing a repetition is necessarily contrastive in the S&C sense : as 5 is a prime number, there is no matricial structure that can support a system of implications on the 5th element[1]. Therefore, if we consider the projective implication triggered by the first 4 elements in a pentadic sequence, we get :

either $\quad a + f(a) + f^2(a) + f^3(a) \xrightarrow{implies} f^4(a)$

or $\quad a + f(a) + g(a) + g(f(a)) \xrightarrow{implies} nil \qquad$ (if $g \neq f^2$)

As a consequence, extensions of plain (square) primary patterns such as $[aaaac]$, $[aabbc]$ and $[ababc]$ (with $c \neq a, b$) all form contrastive sequences. Moreover, the redundancy existing within the first 4 elements opens up to the possibility for another contrast (denoted x) to develop in 4th position, without compromising the detection of the carrier system. This leads to additional types of contrastive sequences : $[aaaxc]$, $[aabxc]$ and $[abaxc]$.

Considering now the possibility that $x = a, b, c$, we obtain 12 prototypes of pentadic configurations, plus 12 more, if we enable $c = a, b$ (see Table 5).

Extensions of aaa	**Extensions of aab**	**Extensions of aba**
$aaaaa$	$---$	$---$
$aaaac$	$aabbc \quad (aabba, aabbb)$	$ababc \quad (ababa, ababb)$
$aaaxc$	$aabxc \quad (aabxa, aabxb)$ $aabac \quad (aabaa, aabab)$	$abaxc \quad (abaxa, abaxb)$ $abaac \quad (abaaa, abaab)$
$aaacc$	$aabcc$	$abacc$

Table 5 : Prototypic pentadic configurations

[1] Unless the expectation of a 6th element is also assumed (see hexadic systems).

A noteworthy point : pentadic configurations from Table 5 can also be viewed (and reparsed) as square sub-system(s), or *stem(s)*, augmented by the insertion of an additional element (an *affix*). For instance, $[aabbc]$ can be envisioned as an infixed b in an $[aabc]$ square stem, while $[abacc]$ appears as $[abac]$ with a suffixed c (a common way to reinforce the closure). Along the same lines, $[aabac]$ can be approached as a "prefixed" version of $[abac]$.

As illustrated on Figure 8, these situations can be modeled by the introduction of a third syntagmatic function h which connects the affix to its most closely related neighbor.

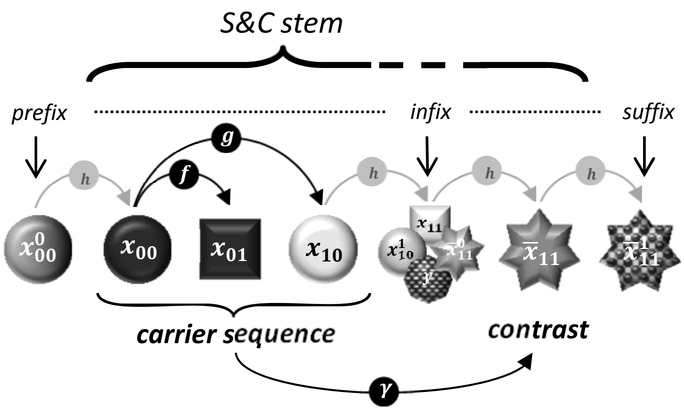
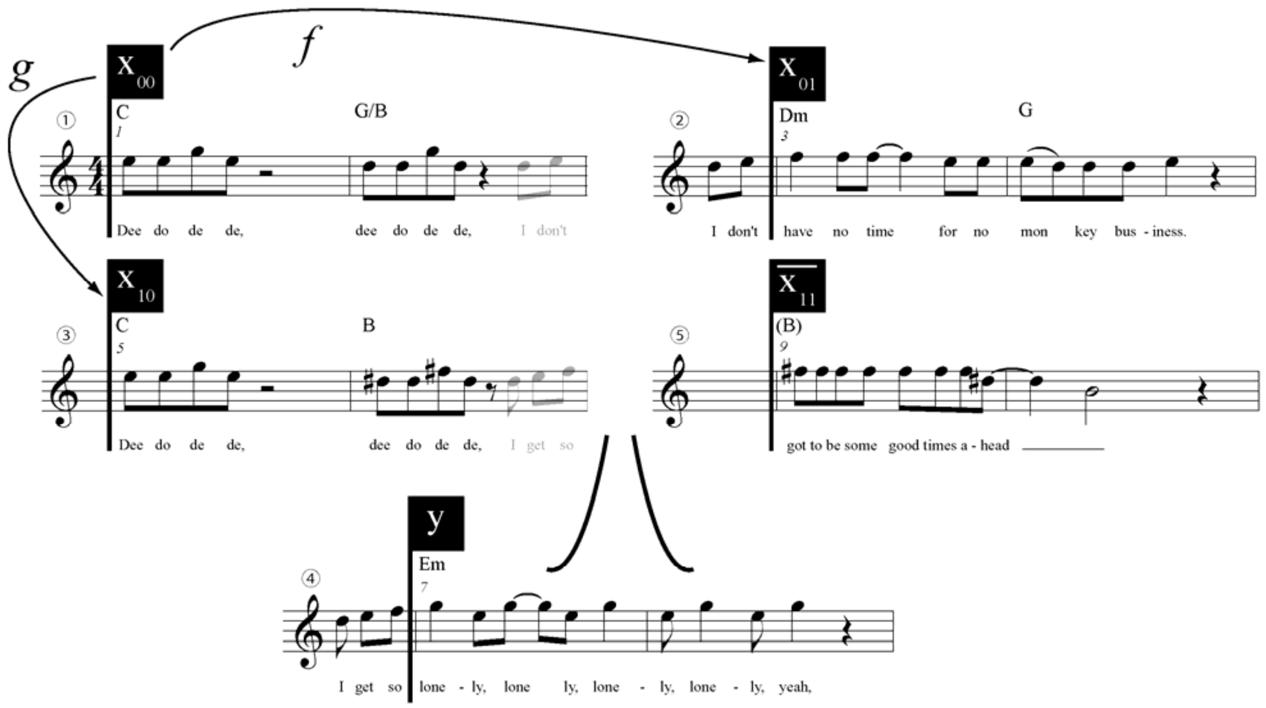


Figure 8 : Synthetic view of several possible pentadic configurations (unfolded forms) based on a square S&C stem $x_{00}\ x_{01}\ x_{10}\ \bar{x}_{11}$

Example n°7

Freddie Mercury – Living on my own
Living on my Own, Parlophone, 1993. Timing : 1'22 – 1'41
"The Freddy Mercury album", International Music Publications Ltd, 1993



Example 7 : this musical segment is excerpted from a song by Freddie Mercury originally released in 1985 in the album "Mr Bad Guy" and remixed in 1993 as a dance version, under which it became famous. This example illustrates a pentadic system, denoted here as $\mathrm{X}_{00}\ \mathrm{X}_{01}\ \mathrm{X}_{10}\ \mathrm{Y}\ \overline{\mathrm{X}}_{11}$, i.e. as a sequence that can be viewed as a square S&C with an extraneous infix.

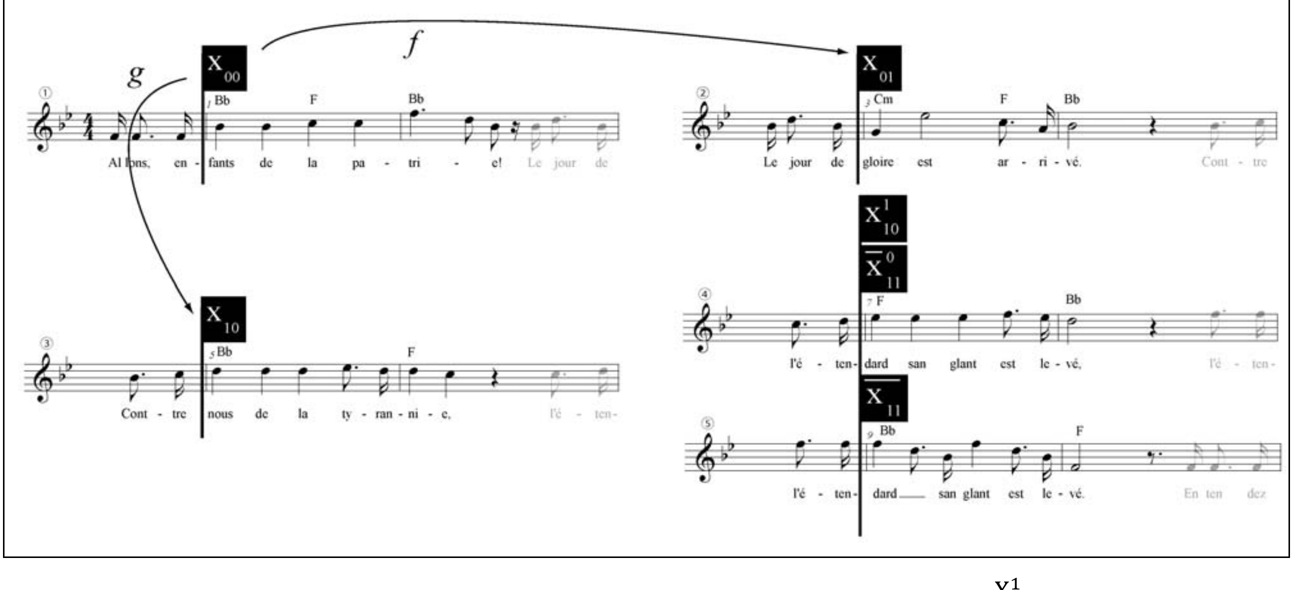
After the exposition of the primer X_{00}, function *f* formulates the introduction of a new element X_{01}. Function *g* relates the primer to a third element X_{10} which *starts-(a)like*, thus inducing a period-like pattern. The first bars in both elements are indeed identical, whereas the second bars can be deduced from each other by a chromatic transposition, creating a tonality shift acting as a transition towards the following element (and, incidentally, a distant sub-S&C at the one-bar scale…).

The fourth element (labeled as Y) could be the contrast of the system, with repetitions, syncopated patterns and a steady tonality of E minor, none of these properties being observed in the previous elements. However, the fifth element, $\overline{\mathrm{X}}_{11}$ concludes the system by returning to non-syncopated rhythmic patterns, in a tonality which is in line with that of X_{10}. It stands as a valid contrast in a square stem $\mathrm{X}_{00}\ \mathrm{X}_{01}\ \mathrm{X}_{10}\ \overline{\mathrm{X}}_{11}$, from which Y would be removed.

According to this description, the inner structure of the segment can thus be written as $aba'xc$, with c representing the final contrast and x an additional extraneous element weakly related to X_{10} and $\overline{\mathrm{X}}_{11}$[1]. Note however that an alternative description could consider that bars 7-10 form the contrast of the segment, namely a double-size segment with harmonic deceleration (these properties constituting a factor of contrast)[2].

Example n°8

Claude-Joseph Rouget de Lisle – La Marseillaise (National French Anthem)
Adapted from l'« Hymne des Marseillais », 1792. Bars 1-10. *Transcribed by ear*



The passage of example 8 is another example of pentadic system, written here as $\mathrm{X}_{00}\ \mathrm{X}_{01}\ \mathrm{X}_{10}\ \genfrac{}{}{0pt}{}{\mathrm{X}^{1}_{10}}{\overline{\mathrm{X}}^{0}_{11}}\ \overline{\mathrm{X}}_{11}$.

[1] It is worthwhile noting that the infix Y is actually re-used, in a different context, at the end of another song of the same album ("I was born to love you", ~ 4'30"), which further supports the hypothesis of an exogenous (or at least detachable) 4th element.

[2] It may happen that several stems are conceivable, or several interpretations of the infix are possible (see also example 10). Rather than a flaw, this should be considered as a feature that reinforces the consistency of the sequence as a pentadic system, as *several descriptions* are able to account for the observed construction and therefore concur to the *system cohesion*.

The 4th element can be viewed as forming (i) partly a sub-system with the 3rd element X_{10} and (ii) partly a sub-system with the 5th element $\overline{\mathrm{X}}_{11}$, depending on which musical dimensions are considered. The following patterns can be observed :

rhymes	$ababb$
harmony	$abcc^{*}c$
melodic contour	$ab\tilde{a}^{0}\tilde{a}^{1}c$
rhythm of the melodic lead	$aa'bb'c$

The 5 elements form a S&C, where the 4th element relates equivocally to both its neighbors. However, the sequence $\mathrm{X}_{00}\ \mathrm{X}_{01}\ \mathrm{X}_{10}\ \overline{\mathrm{X}}_{11}$ forms a valid square stem. Here again, the segment can be considered as a pentadic system but an alternative would be to consider that the last 4 bars form the contrast, on the basis of an oversized last element.

Generally, pentadic systems tend to require a more in-depth analysis of musical material than square systems, so as to deal with their unevenness within the S&C framework. Identifying which of the 5 elements plays the role of the affix may lead to several concurrent hypotheses between which it is not always possible to arbitrate (unless a shortened version of the same system is observed somewhere else in the piece). However, arbitration may not be needed : the very existence of multiple hypotheses is *per se* a strong evidence of the 5 elements being part of a same segment.

Hexadic systems

Hexadic systems result from projective implications across sequences of 6 elements, and can be approached as *rectangular* (rather than square) S&Cs. Two main categories can be distinguished (*tall* and *wide*), depending on the span and the interactions of the syntagmatic relations :

$$[a+f(a)] \ + \ \left[g_1(a)+g_1\big(f(a)\big)\right] \ + \ g_2(a) \xrightarrow{implies} g_2\big(f(a)\big) \qquad \text{(Tall hexadic)}$$

$$[a+f_1(a)+\ f_2(a)] \ + \ g(a) \xrightarrow{implies} g\big(f_1(a)\big)+g\big(f_2(a)\big) \qquad \text{(Wide hexadic)}$$

Table 6 inventories typical hexadic patterns, based on three carrier elements (*a, b, c*) and on 1 or 2 contrasts (noted as horizontal bars). Note that some hexadic sequences are better explained as square S&Cs with a nested subsystem (3rd line in the table) and tall hexadic S&Cs with double contrasts can be ambiguous with a square S&C followed by an independent dyad (4th line)[1].

<table>
<tr><th>Configuration</th><th colspan="2">Plain forms</th><th colspan="2">Contrastive forms</th></tr>
<tr><td>Tall hexadic S&C</td><td colspan="2">$aabbcc$
$aaaabb$
$ababab$</td><td colspan="2">$aabbc\bar{c}$
$aaaab\bar{b}$
$ababa\bar{b}, abab\bar{a}b, abab\overline{ab}$</td></tr>
<tr><td>Wide hexadic S&C</td><td colspan="2">$abcabc$
$aabaab$
$aaabbb$</td><td colspan="2">$abcab\bar{c}, abca\bar{b}c, abca\overline{bc}$
$aabaa\bar{b}, aaba\bar{a}b, aaba\overline{ab}$
$aaabb\bar{b}, aaab\bar{b}b, aaab\overline{bb}$</td></tr>
<tr><td>Square S&C with nested dyad</td><td>$aba(x_0x_1)b$
$ab(x_0x_1)ab$</td><td>$aab(x_0x_1)b$
$aa(x_0x_1)bb$</td><td>$aba(x_0x_1)\bar{b}$
$ab(x_0x_1)a\bar{b}$</td><td>$aab(x_0x_1)\bar{b}$
$aa(x_0x_1)b\bar{b}$</td></tr>
<tr><td>Plain S&C followed by dyad</td><td colspan="4">$(aaaa)(yz) \quad (abab)(yz) \quad (aabb)(yz)$</td></tr>
</table>

Table 6 : Typical patterns of hexadic sequences

[1] Iterative hexadic systems also exist, based on sequential iterations of 2 syntagmatic functions, but they boil down to triadic systems at the immediately upper scale. For instance :

$$[a+f(a)] \ + \ \left[g(a)+g\big(f(a)\big)\right] \xrightarrow{implies} \left[g^2(a)+\ g^2\big(f(a)\big)\right]$$

$$[a+f(a)] \ + \ \left[g(a)+g\big(f^2(a)\big)\right] \xrightarrow{implies} \left[g^2(a)+g^2\big(f^3(a)\big)\right]$$

In Example 9, the first 6 bars of the first movement of the "Autumn" concerto by Vivaldi is an example of an (almost) plain "wide" hexadic system $[aabaa\bar{b}]$, where function g combines a change of dynamics and the transposition of three of the four voices to the lower octave. Here, the contrast γ is almost "identity", except for the final part of the accompaniment of X_{12}, where, on the score, function g stops applying to one of the voices.

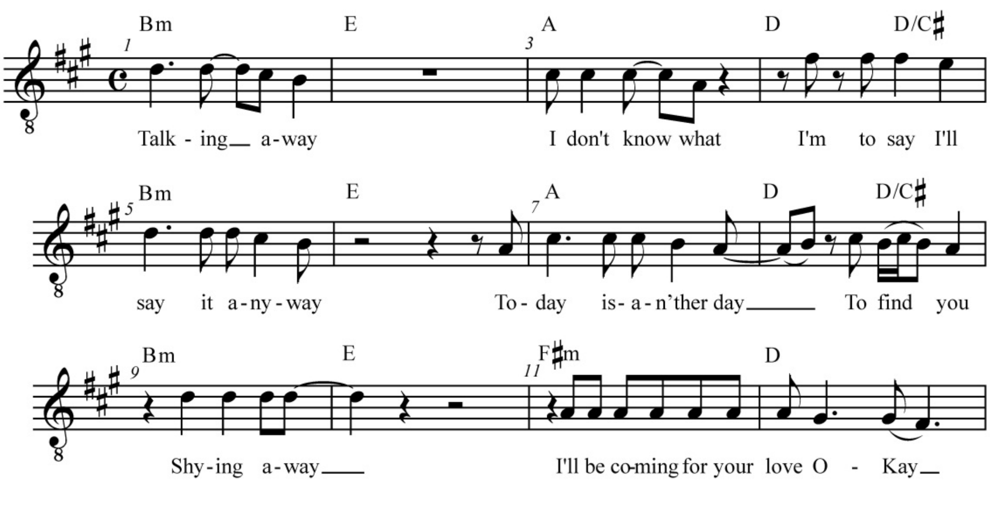
Example n°9

Antonio Vivaldi – Concerto n°3 in F Major ("Autumn")
Op. 8, Ryom Verzeichnis 293, 1st mvt, 1723. Bars 1-6
Reduction from a transcription by H. Sawano (http://sound.jp/kazane)

Example 10 is a transcription of the first verse of a song by the Norwegian pop group A-Ha, which illustrates a "tall" hexadic segment essentially based on a harmonic system behaving as $[ababa\bar{b}]$ and a melodic line sounding like $[ab\tilde{a}b'a'\bar{b}]$. This segment can also be viewed as a triadic system $[aaa']$ on the harmonic dimension, at the immediately upper scale.

Example n°10

A-Ha – Take on me
Composers : M. Furuholmen, M. Harket, P. Waaktaar-Savoy
Label : Warner Bros. Records (1985). *Verse 1. Transcribed by ear*



Larger systems

In line with the same logic, heptadic (7-element) segments can generally be approached as various affixed versions of square or hexadic systems, where the insertion of musical material tends to suspend the implication process. We leave it to the reader to envision constructions such as : $[abcabx\bar{c}]$, $\left[ababax\bar{b}\right]$, $\left[aba(x_0x_1)\overline{bb}\right]$, etc…

Similarly, some 8-element segments can be described as complex irregular forms deriving from smaller carrier systems. But some may also result from *cubic* systems $(2 \times 2 \times 2)$, i.e. systems based on properties evolving along three distinct musical dimensions and/or variation cycles (Deruty, Bimbot & Van Wymeersch, 2013). In general, however, cubic systems can be approximated as two successive square systems, by neglecting the syntagmatic function with the longest span. Alternatively, they can be approached as a single square system at the immediately upper time-scale, by grouping neighboring elements two by two.

Nonadic systems (3×3) are perfectly conceivable, but they turn out to be quite rare in pop music.

Tiling

Occasionally, 2 successive systems overlap, thus creating a situation where the contrast of a given S&C and the primer of the next S&C are either superposed (played/heard at the same time) or merged into a single element, thus functionally acting both as a contrast and a primer. This configuration corresponds to "grouping overlaps and elisions" as defined by Lerdahl and Jackendoff (1983).

S&Cs at different timescales

S&Cs detectable around a given timescale usually exhibit nested *sub-systems* at lower timescales, and they may themselves be embedded in larger *super-systems*. When working at a given timescale, subsystems are useful to comfort the consistency of a segmentation hypothesis (for example, the succession of two "sentences" starting by the same presentation, but two distinct continuations, constitutes a "period" at the immediately upper scale).

In Example 11, this transcribed segment from a recent American pop song exhibits a typical period-like form $[abac]$ at the chosen timescale and granularity (4 × 2 bars). Function g can be considered as "*identity*". Function f is a relatively straightforward transformation on the first half of the primer, but a more complex one of its second half ($b \approx a_1'$).

In fact, the sequence X_{00} X_{01} can itself be viewed as a smaller S&C at the immediately lower timescale and granularity (4 × 1 bar), where tracks 1, 3 and 4 form a contrastive system (while tracks 2 and 5 form a plain system). At an even lower time-scale (4 × 1/2 bar), track 1 of segment X_{00} exhibits a sentence-like pattern, while tracks 2, 3 and 5 show a period-like organization and track 4, a progression.

This example illustrates how the S&C model is able to account for the inner organization of musical segments at several time-scales simultaneously : the density of such multi-scale relationships within a musical segment (and especially those which involve the primer) can be considered as an indication of its structural consistency (Deruty, Bimbot & Van Wymeersch, 2013).

Altogether, a consistent structural segment in the framework of the S&C model fits well with the general definition given by Spencer & Temko (1988, p. 31) : "*a major structural unit perceived as the result of the coincidence of relatively large numbers of structural phenomena*".

Example n°11

Britney Spears – Heaven on Earth
Composers : Mc Groarty, Huntington, Morier
Blackout, Jive, 2007. Timing : 1'23 – 1'39. *Transcribed by ear*

Scope and Limits of the S&C Model

The S&C model is particularly suited to musical styles where the normative phrase structure involves the presentation of 2 bars units in an 8-bar framework. The square S&C scheme works very well for pop music (all the more since this fits standard musical verse schemas), as well as for some dance-based genres, as one may come across for instance in Baroque suites.

There is also a direct connection between the prototypic "square" form of the S&C model and the concept of *carrure* originally pointed out by Fetis (1830, pp. 60-62), and extensively used by Mozart and many other composers after him. According to (Brennet, 1926) the *carrure* (literally meaning the "build" of a person, in French) is defined as *the symmetry established between portions of the musical phrase, so as to divide it into fragments of equal length. The term "square" is specifically used for melodic forms whose periods proceed by 4 or multiples of 4*.

In this section, we explore, on two examples from the classical genre, how the use of the S&C model may inform conventional music analysis and we discuss the S&C model as an interesting way to apprehend musical segments as *computational objects*.

Formal types

When considering the structure of musical segments, Schönberg (1967, pp. 21-30), and after him Caplin (1998), define two types of inner organizations, referred to (by Caplin, pp. 9-12) as *formal types* : the *period* and the *sentence*. Both types are normatively 8-bar segments, even though they may last for 16 or even 32 bars. They begin with what Schönberg calls a "two-measure phrase" (Caplin, a "two-measure idea", or a "basic idea"), which occupies the first quarter of the segment (typically what we consider as the primer).

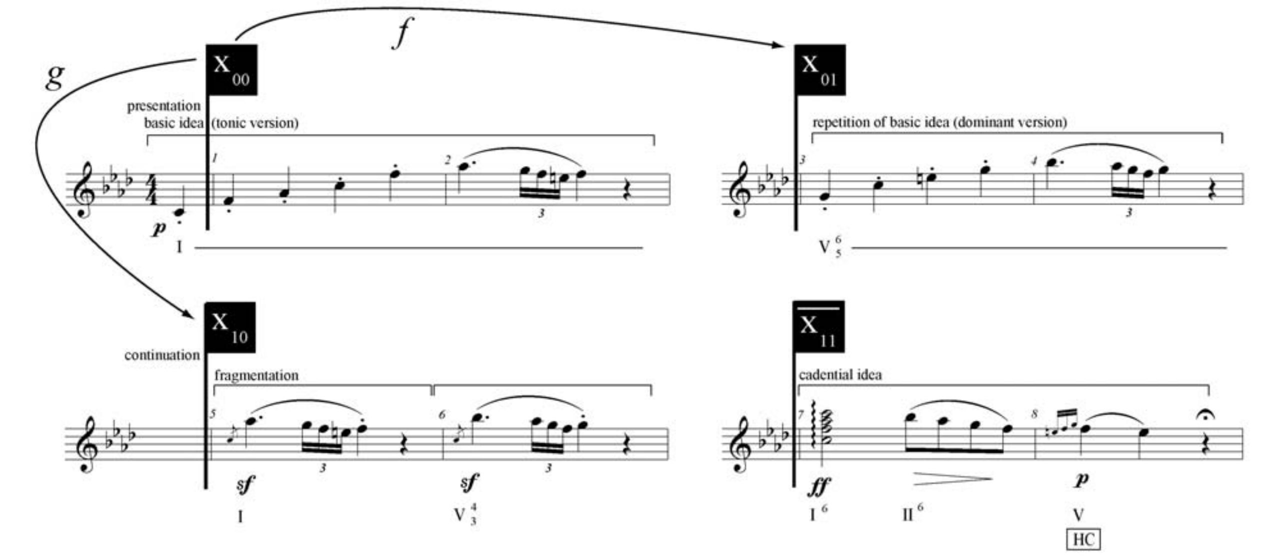
In sentences, the basic idea is repeated immediately, forming what Caplin calls the *presentation*. The second part of the sentence, the *continuation*, can either be the result of transformations (or formal processes)[1] of the presentation, or the introduction of new ideas. As a general form, the sentence may be written as $[aabc]$, with both b and c being unspecified.

The period differs from the sentence in the postponement of the repetition. This is done using the introduction of what Caplin calls a "contrasting idea[2]" between the two occurrences of the basic idea, which normatively lasts a quarter of a period. The first half of the period is called the *antecedent*, and the second half the *consequent*. A period may therefore be written as $[abac]$, with c being unspecified.

According to Caplin (1998, pp. 9-10), Example 12 *"presents perhaps the most archetypical manifestation of the sentence form in the entire classical repertory"*. The *"basic idea"* takes place in measures 1-2 and is repeated as a dominant version in measures 3-4. The extract proceeds with the *"continuation"*. Here, measures 5-6 show elements derived from the basic idea by means of *"fragmentation and harmonic acceleration"*, and finally, measures 7-8 present the *"cadential idea",* which effects closure for the entire segment, using conventional harmonic and melodic formulas.

Example n°12

Ludwig van Beethoven – Piano Sonata in F Minor
Op. 2, 1st mvt, 1795. Bars 1-8
Adapted from (Caplin, 1998), p. 10



[1] This concept resonates well with the concept of syntagmatic relation introduced in this article.

[2] In the present article, the "*contrasting idea*" of (Caplin, 1998) corresponds to our concept of *opposition* (i.e. something "different") but does not identify with that of *contrast* (something that departs from a logical system).

In the framework of the System & Contrast model, it can be observed that :

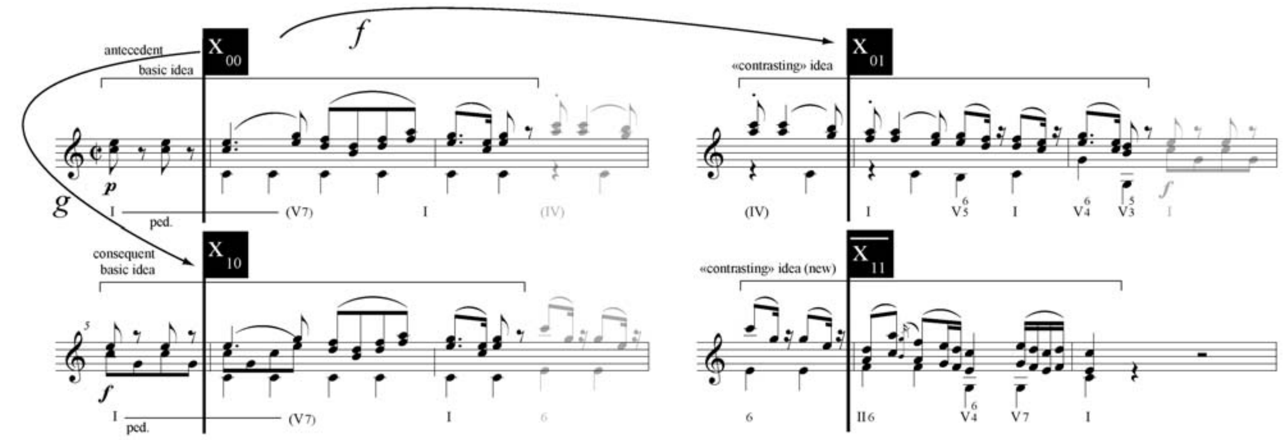
- The <u>note duration values</u> follow an $[aabc]$ pattern over the passage.
- The <u>intensity level</u>, as indicated on the score, also matches a clear $[aabc]$ pattern.
- Even though the melodic lines of X_{00} and X_{01} are not exactly alike, they relate to each other as a diatonic (or tonal) transposition, the melody of the whole segment forming an $[a^0 a^1 bc]$ pattern.
- The <u>chord sequence</u> tends to follow an $[aba'c]$ pattern[1] and its <u>rate of variation</u> follows an $[aabc]$ pattern over the passage, with the harmonic acceleration contributing to the closural effect.

In summary, several musical dimensions in this passage follow a well-identifiable S&C pattern, in line with the predominant impression towards a sentence form. The relationships between measures 5 and 6 in X_{10} and measures 2 in X_{00} and 4 in X_{01} (fragmentation) are somehow overlooked by the above description, even though they undoubtedly contribute to the cohesion (and to the closure) of the segment, at a lower granularity ($a^0 + \tilde{a}^1 + \left[\frac{a^0}{2} + \frac{\tilde{a}^1}{2}\right] \nrightarrow c$)[2].

Let us consider now Example 13, also chosen by Caplin (1998, p. 12) as an archetype of the *period* form. The first element is the *basic idea*. The second element is the presentation of new material and ends, here, with a weak cadential formula. The third element is a re-exposition of the basic idea. The fourth element is an opposition to the basic idea X_{00} (but also to the "contrasting idea" X_{01}), and it ends with a strong cadential formula.

Example n°13

Wolfgang Amadeus Mozart – Eine kleine Nachtmusik
Köchel 525, 2nd mvt, 1787. Bars 1-8
Adapted from (Caplin, 1998), p. 12



In this example, all the major musical dimensions form primarily an $[abac]$ pattern, except for minor modifications between X_{00} and X_{10}. The rhythmic pattern of a, b and c are all different. But, whereas X_{00} and X_{01} are somehow off-beat in their first bar, they re-synchronize on an identical rhythmic pattern in their second bar. This is not the case for the contrast $\overline{\mathrm{X}}_{11}$, whose melodic rhythm is in total disparity with all the other elements in the segment. Note also the existence of some parallelism between the harmonic placements in X_{01} and $\overline{\mathrm{X}}_{11}$ (harmonic rhythm), which reinforces the contrastive

[1] The a' element is also some kind of b'', where b'' would mean « ends-like » b.

[2] Note also, as one of our reviewers pointed out, that measures 5-8 can be seen as a nested S&C (at the one-bar level) with measure 7 appearing as a diatonically transposed expansion of measure 5. This is also supported by the dynamics of the 4 measures forming a sentence-like pattern : [*sf sf ff p*].

effect of their inversion at their termination (resp. I-V versus V-I), all the more since there is no such inversion between X_{00} and X_{10}.

Projective implications versus cultural conventions

In direct lineage with Narmour's I-R model and Cognitive Rule Mapping scheme, the S&C model relies essentially on the analysis of internal relationships, patterns and projective implications within music segments, i.e. an *intra-opus* point of view. The question may arise on how this view articulates (and possibly conflicts) with analyses based on *inter-opus* stylistic similarities, i.e. the recognition of "style structures" based on learnt conventions, and in particular, solidified closed forms such as phrases and cadence types.

Narmour (2000) raises this point (p. 332) and stresses that "*even though […] stylistic repetition and iterative rule appear so interconnected as to be difficult to tease apart, style does not circumscribe rule – and vice versa"*. Similar considerations are also found in the field of Engineering Sciences, in the articulation between Shannon information (Shannon, 1948) and Kolmogorov complexity (Kolmogorov, 1963), which lead to the definition of two types of information, themselves related to two data encoding strategies (see for instance, Grünwald, 2010).

"In the Shannon approach, the method of encoding objects is based on the presupposition that the objects to be encoded are outcomes of a known random source"(Grünwald, 2010). Shannon information relates to uncertainty, by estimating the probability of an object as its *statistical prevalence* within a general population of possible events stemming from a probability distribution. In short, Shannon information measures how many bits are necessary to single out the observed object from a predefined general population.

"In the Kolmogorov complexity approach, […] the encoding of an object [i.e. its compressed version] is a short computer program that generates it" (Grünwald, 2010). In that context, Kolmogorov information theory estimates the probability of the object as a function of its *internal redundancy*, via the number of bits required to specify a compression program that generates the observed object.

Quoting Grünwald (2010) again : *"Shannon ignores the object itself but considers only the characteristics of the random source of which the object is one of the possible outcomes, while Kolmogorov considers only the object itself to determine [its] compressed version irrespective of the manner in which the object arose"*. In short, Shannon's approach to information is essentially statistical (and knowledge-based) while Kolmogorov complexity is computational (and observation-driven), resulting in a duality between the two concepts[1].

This duality is reflected in the roles played by cultural conventions and projective implications in the perception of music : "*difficult to tease apart*" but "*not circumscribing each other*", as Narmour says. In music genres for which conventions are well-defined, strong and stable, reliable underlying probability distributions are conceivable as a good support to Shannon-based paradigms, i.e. qualifying musical material on the basis of inter-opus analogies. However, when dealing with less formal, more intuitive, versatile and rapidly evolving types of music (as is arguably the case for nowadays pop music), the reliability or even the existence of such probability distributions is questionable. In this case, the Kolmogorov-based approach may appear as a valuable alternative and/or a complementary strategy to describe and characterize some structural aspects of music on the basis of intra-opus properties, without requiring an explicit formulation of stylistic and idiomatic specificities of each and every music genre.

[1] Even though some theoretical equivalences have recently been established, but under specific assumptions (Li & Vitanyi, 2008)

In any case, music is not a “uniquely decodable code” : multiple yet undecidable hypotheses and explanations may coexist, even within each viewpoint, as an entire part of the artistic dimension of music.

Limitations of the S&C model

In its current form, the S&C model covers particularly well a wide variety of forms in modern popular western music styles and it can handle a number of forms observable in music from the late baroque, classical, romantic eras, thus providing a quite generic framework to a broad range of situations.

However, the S&C fails to account for more sophisticated art music structures found in some classical pieces, as well as in modern, 20th century, or contemporary music, whose constructions can depart radically from intuitive similarities[1]. Moreover, the S&C model does not cover all aspects of form, in the many dimensions along which it can be approached, in particular, structural “functions”, in the sense of Spencer & Temko (1988).

Here, we illustrate an example of the limits of the S&C with the passage from Wagner’s Parsifal (Example 14) : these 16 bars cannot be reasonably described (at any conceivable mid-level time-scale) with the System & Contrast model. It is indeed striking how, in this musical flow, it is virtually impossible to identify elements that would relate to each other on any musical dimension to form clear patterns. If this example were to be viewed as an instance of “infinite” or “unending” melody, it appears as music based on permanent novelty and seems to have been designed to exhibit very little redundancy on the morpho-syntagmatic level.

Conversely, for “patterned” music (especially pop music), the potential of the S&C model lies in its *polymorphism*, i.e. its capacity to encompass under a single formalism all the musical dimensions that create patterns, independently of their role or value in the musical discourse. However, this versatility requires an additional step in the discovery process : that of identifying the relevant musical dimensions and codifying their relationships within a potentially very large set of musical parameters and factors, which, in some situations, may turn out to be operationally complex.

At this point, the scope of the S&C model is strictly that of a *descriptive data model*, accounting for specific organizational aspects induced by similarity in music. It can feed and fuel compositional, perceptual or functional analyses of a musical passage but it certainly does not replace them.

[1] This remark is not a matter of disqualification of the model but rather as an encouragement to explore it for a large range of music genres, though with discernment : its success or difficulty in handling such or such passage or type of music is *per se* informative.

Example n°14

Richard Wagner – Gurnemanz' Monologue
Parsifal (Act I). Wagner Werk Verzeichnis 111, 1882
"Parsifal, Klavierauszug zu zwei Händen", Ed. R. Kleinmidjel, p. 19, 1911

S&C Discovery as a Model Matching Task

Grouping and parsing

The decomposition of a passage in terms of S&Cs can be approached as a grouping process (Lerdahl & Jackendoff, 1983)[1], where the grouping criterion results from the possibility to identify, in the observed data, patterns that match the S&C data model : the description of a musical segment is viewed as an information coding scheme and the grouping operation therefore aims at seeking, as parsing hypotheses, the S&C decompositions which provide the most "economical" descriptions of the passage.

In this section, we introduce principles, criteria and methods to this aim, which highlight the essential role played by the determination of the <u>primer</u>, in the process of S&C discovery.

As general guidelines, we refer to the following hypotheses :

- In a structural segment, the various musical dimensions may not follow the *same* structural pattern (some may for instance behave as $[abac]$, others as $[aabc]$) but all these patterns are assumed to contribute to the S&C, provided they exhibit a <u>synchronous</u> (or congruent) <u>behavior</u>.
- In general, *not all musical dimensions* do participate to the S&C, but only a subset of them, while the other musical dimensions exhibit unstructured behavior. This requires a stage of <u>selection of variables</u>.

---

[1] Note however that, whereas *grouping structure* relies on a tree-based hierarchy driven by affinities between strictly adjacent segments, the S&C model assumes a *matrix scheme* which is able to account for *associational* structure (Lerdahl & Jackendoff, 1983), i.e. relations between elements which may not be immediately contiguous.

- Segmentation of a passage using the S&C model results from a joint optimization among concurrent S&C hypotheses over the whole passage.
- At the working timescale, system size is assumed to range between 3 and 7 morphological elements, while smaller (resp. larger) segments are considered to relate to lower (resp. higher) time scales.

Simplicity principle and law of parsimony

Following Narmour's conception of cognitive rule-mapping, humans tend to achieve intuitively the joint estimation of structuring properties and their relationships by eliciting, between several possibilities, the one(s) which seem(s) the most salient[1]. Such a "cognitive" strategy relates to the Ockham's razor principle (also called the *law of parsimony*) (Ockham, 1323) which assumes that, among several possible ways to describe a same set of observations, the preference tends towards the simplest explanation. Recent research in experimental psychology also hypothesizes "*the search for simplicity [as] a fundamental cognitive principle*" (see for instance, Chater 1999).

In the field of Engineering Sciences and Information Theory, the Minimum Description Length (MDL) approach (Rissanen, 1978) is a particular instance of this principle, based on the fundamental idea that "*any regularity in a given set of data can be used to compress the data, i.e. to describe it using fewer symbols than needed to describe the data literally*" (Grünwald, 1998).

Once again, the upper compression bound of a set of data corresponds to their Kolmogorov complexity (Kolmogorov, 1963), i.e. the size of the shortest *program* that outputs the data. The MDL criterion restricts the search of the optimal compression scheme to a subset of allowed *codes*, called the model class, which is assumed to be reasonably efficient, whatever the data at hand.

Based on these principles, the adequacy of a S&C in describing a musical passage can be formulated as an information-theoretic criterion which expresses how much compression gain can be achieved by modeling the passage as a S&C rather than describing it literally.

S&C model matching by MDL criterion

In this context, the musical passage $S = x_{00}\, x_{01}\, x_{10}\, \bar{x}_{11}$ plays the role of the literal data, the S&C framework is the model class (the "coding scheme") and the S&C description of the passage, $M = (x_{00}, f, g, \gamma)$ is the compressed data.

Let us assume that the quantity of information (i.e. the number of bits) which is required to describe S literally writes :

$$q_0(S) = q_0(x_{00}) + q_0(x_{01}) + q_0(x_{10}) + q_0(\bar{x}_{11})$$

where $q_0(x_{ij})$ denotes the quantity of information required to describe each element of the system. When compressed as M, the quantity of information needed to describe S becomes :

$$q_M(S) = q_0(x_{00}) + q_M(f) + q_M(g) + q_M(\gamma)$$

$q_M(\theta)$ denoting the quantity of information needed to encode a given mapping function θ.

The compression gain achieved by the S&C model can therefore be expressed as the difference between these two costs, namely :

$$\Delta q(S|M) = q_0(S) - q_M(S)$$

For a given sequence S, the MDL description M^* is therefore chosen as the one which minimizes $q_M(S)$, i.e. maximizes $\Delta q(S|M)$.

[1] Naturally, there can be cases for which different sorts of similarities are comparably "salient", and saliency may vary over people.

It is not within the scope of this article to develop explicit expressions of q_0 and q_M. In common music experience, the MDL description M^* is estimated intuitively by the listener on the basis of the most apparent similarities between morphological elements. Once normalized between 0 and 1, empirical values for q_M should more or less behave as in Table 7.

Type of similarity	a	$\tilde{a}, a^k$	$f(a), g(a)$	a'	b
q_M	0	$\varepsilon \ll 1$	$\alpha, \beta \leq \frac{1}{2}$	$\frac{1}{2}$	1

Table 7 : Empirical range of normalized values of q_M, for various types of mapping functions

As f, g (and indirectly γ) are defined in reference to the primer x_{00}, this element plays a key role in the compression criterion : not by its intrinsic (literal) complexity, but depending on how much the *forthcoming elements* in the system can be economically explained on its basis[1]. Identifying optimal primers is therefore essential to the parsing process.

Selection of variables

Jointly to the estimation of the syntagmatic relations, S&C MDL model matching operates a *selection of variables* over which the most compact description can be achieved.

Let's denote as $\mathbb{S}$, the subset of structuring properties of segment S and $\overline{\mathbb{S}}$ all the other (non-structuring) properties. The total quantity of information q_M decomposes into two terms, corresponding to the quantity of information required to encode S in each separate subspace :

$$q_M(S) = q_M(S|_{\mathbb{S}}) + q_M(S|_{\overline{\mathbb{S}}})$$

While the first term in the sum, is smaller than its literal counterpart $q_0(S|_{\mathbb{S}})$ because of the existence of simple syntagmatic relationships within $\mathbb{S}$, the second term is merely invariant, as no compression can be achieved from variables which vary unrelatedly with the model.

As a consequence, the compression gain actually writes as :

$$\Delta q(S|M) = q_0(S) - q_M(S|_{\mathbb{S}})$$

and its decrease is entirely attributable to the variables living in subspace $\mathbb{S}$. Optimizing the match of a segment to a S&C model provides simultaneously an estimation of $\mathbb{S}$, as the subset of variables responsible for the maximum compression gain.

In general, not all musical dimensions form S&C patterns in a structural segment, but only a subset of them, while the other ones vary in a non-organized way. For instance, the lead part played by the central instrument during a solo may temporarily break free from any structured patterns and the organization of the segment may be only governed (weakly) by the underlying harmonic progression. In many rap pieces, the vocal lead is unstructured, neither in terms of melody, nor harmony, and the structural "clock" is enforced by the background loop, etc… Note also that musical dimensions which follow S&C patterns often vary across successive structural segments.

Segmentation into S&Cs

When parsing morphological elements into structural segments across a passage or an entire music piece, the *boundaries* of the successive S&C need to be optimized *jointly* with the S&C themselves (Bimbot, Sargent, Deruty, Guichaoua & Vincent, 2014). Valid segmentation hypotheses result from a simultaneous identification of the contrast of the finishing segment and of the primer of the forthcoming one.

[1] Somehow in analogy with the root of a chord.

As a concrete illustration, let us consider Example 15.

Example n°15

Loreen – Euphoria
Composers : T. G:son & P. Boström. Label : Warner Music
Eurovision Song Contest winner 2012 (Sweden). *Transcribed by ear*

In the first 1'20" of this trance-inspired euro-pop song, the inner structure of each of the first two verses (bars 1-8 and 9-16) clearly forms a dominant period-like pattern $[abac]$ (supported by a plain $[abab]$ rhyme system). Then, the chorus proceeds to a sudden change of *regime* : the harmony becomes $[abab]$ while the melodic system exhibits a dominant $abb'c$ pattern[1], prone to create a structural uncertainty on the status of bars 17-18 ("Euphoria") as a primer.

This ambiguity is however resolved by at least two other factors :

- Bars 23-24 clearly create a contrast with the 3 previous morphological elements (17-22), forming a very distinctive "b" in an $aaab$ pattern in terms of (i) "on-the-beat" vs "off-beat" note placement, and (ii) "rising" vs "falling" melodic termination. In particular, the insistent and syncopated repetition of "up, up, up…" over 1½ bar constitutes a definite "punctuation mark" in this context, in contrast with the smoother flow of the previous bars.
- Hypothesizing a structural segment which would start on bar 19 (rather than 17) would create a very inconsistent organization for the *second part* of the chorus (4th line of the transcription), given that bars 33-34 necessarily constitute the primer of a new occurrence of the song's verse (5th line).

This example illustrates how competing hypotheses can be arbitrated by means of the empirical minimization of an overall description cost, which elicits the most parsimonious explanation. The preference for a segment boundary arises from the detection of a plausible contrast *before* the boundary but, first and foremost, from the identification a reasonable primer

[1] Exactly $abbc$ for what concerns the rhymes

just *after* the boundary : the primer appears simultaneously as an element which is *poorly suited to the previous S&C* (typically, it is located after the contrast and shows no clear syntagmatic relationship with it) and an element which is *essential* to trigger the forthcoming system in the *next* segment on the basis of compact morpho-syntagmatic relationships

As was noted in (Bimbot, Le Blouch, Sargent & Vincent, 2010) structural segments are perceived as both *autonomous* and *suppressible* within the musical flow because (i) their elements form a consistent set of their own (property of autonomy) and (ii) none of these elements are necessary to explain the elements of neighboring systems (property of suppressibility).

In a vast majority of cases, the joint estimation of successive S&Cs provides disambiguated boundaries and conditions the global segmentation towards uniqueness. Occasionally, local ambiguities cannot be resolved, leading to a very small graph of residual hypotheses.

Communication Functions of the Contrast

Communication is a very broad area which covers a variety of aspects dealing with the exchange of information, messages thoughts and affects, and music is undoubtedly a communication process. In that context, the relationship between information, expectation and surprise in music is a very prolific subject, addressed in a number of research works, including (but not limited to) Meyer (1967), Narmour (1977), Huron (2006), Abdallah & Plumbley (2009), …

The S&C model provides additional perspectives to the question, from the point of view of Communication Sciences : the contrast element can be interpreted as a digital *modulation* of the last element of the *carrier* system. It also appears as some sort of *delimiter* acting as a musical *punctuation mark* within the musical flow and it can be understood as a *prediction residual* creating a *surprise* element which acts as a punch-line.

The contrast as a digital modulation

In the field of electronic engineering and telecommunications, *modulation* is defined as the process of varying one or more properties of a known *carrier signal* (usually a periodic waveform) with an unknown *modulating signal*. The modulation signal conveys information which fluctuates at a lower frequency than that of the carrier.

In analog communications, the modulation is applied continuously to the carrier signal as a function of the information content. In digital communications, the signal is discrete, and modulation consists in changing, at specific instants, the values of the bits constituting the code-words. In both cases, retrieving the modulating information from the composite signal is called *demodulation*.

As pointed out in the introductory part of the S&C, the sequence $S_0 = x_{00}\, x_{01}\, x_{10}\, x_{11}$ can be viewed as a carrier signal, composed of a small number of distinct states for each musical dimension. The fact that the last element of the carrier sequence S_0 is fully predictable (i.e. redundant) provides the opportunity to insert additional information by altering one or several properties (bits of information) of x_{11}. This yields the contrasted sequence $S = x_{00}\, x_{01}\, x_{10}\, \bar{x}_{11}$, where $\bar{x}_{11}$ (and therefore γ) can be understood as a *digital modulation*[1] of the last element of S_0.

In this modulation scheme, the receiver does not know the carrier signal in advance and must reconstruct it (a *blind carrier recovery* problem, in Engineering Sciences). In view of the S&C model, the carrier is reconstructed (at the receiver's end) by inferring the functions f and g from the sequence $x_{00}\, x_{01}\, x_{10}$ (this task being easier if f and g are cognitively simple).

[1] This particular use of the word *modulation* departs from its usual meaning(s) in the field of music.

When the contrastive element of the system occurs ($\bar{x}_{11}$), the receiver is able to identify it as such and to analyze its deviation from the expected carrier element (x_{11}) so as to *demodulate* the contrast ($\bar{x}_{11}\ vs\ x_{11}$)[1].

The contrast as an implicit delimiter

The elements forming human communication messages as well as digital data streams are generally grouped into larger units which constitute blocks of information. In Computer Science, a delimiter is a sequence of one or more symbols used to specify (explicitly or implicitly) boundaries between separate, independent regions in data streams. This is typically the role of punctuation marks in natural human language, which help structuring logically the linguistic message.

Whereas, in written communication, punctuation marks are inserted explicitly in the textual content as specific signs, they take a very different form in oral communication : in spoken language, the logical organization of the discourse is mainly rendered by *prosodic* markers at the end of logical groups of words : introduction of pauses, significant modifications of the intensity, the syllable flow or the vowel durations, inversion of the shape of the intonation contour, etc… Note that these prosodic modifications are realized simultaneously with the articulation of phonemes, i.e. they are embedded within the other levels of the linguistic message.

An interesting parallel can be established between these prosodic processes and the implication patterns occurring in the musical flow as described by the S&C. Indeed, the contrast manifests itself as the last item of a sequence by a significant deviation of its musical properties from their current course. Dually, spoken language uses mechanisms which consists in modulating "musical" dimensions of speech, in order to mark the logical organization of the discourse.

In addition, as for spoken language, the S&C segment boundary falls *after* the contrast which does not only mark the end of the current S&C but also announces the advent of a new one.

The contrast as a thwarted prediction

In the dictionary, the entry *surprise* returns definitions such as "an emotional state experienced as the result of an unexpected event", "the difference between expectations and reality", "the gap between our assumptions and expectations and the way that those events actually turn out" or "the end result of predictions that fail"[2],[3].

Creating surprise is some form of *art*. It plays an important role in novels, plays, jokes, shows, social events,… and of course, in music. Needless to insist on how well the above definitions of *surprise* does indeed apply to that of *contrast* as used in this article.

Generally speaking, the contrast does not constitute a completely novel and unforeseeable event in the musical flow. Firstly, its location is more or less predictable. Secondly, the contrast itself tends to share similarities on some musical dimensions with the carrier system and therefore, it is not totally new in every respect. Therefore, the contrast can be understood as a phase of the musical narration where an unknown quantity of surprise is released, after a sufficient amount of information has been disclosed, so that the contrast gets its full *value*.

Depending on the familiarity of the listener with the musical genre of the piece, the contrast may sound quite surprising or

---

[1] Assuming that the internal organization of structural segments in music is governed by a carrier-modulation process should of course be understood as an abstraction which does not reflect all the aspects at work in the process of musical composition or listening.

[2] Note that surprise can be pleasant, unpleasant or simply neutral.

[3] Surprise must be distinguished from *novelty*, i.e. "the quality of being new, original or unusual".

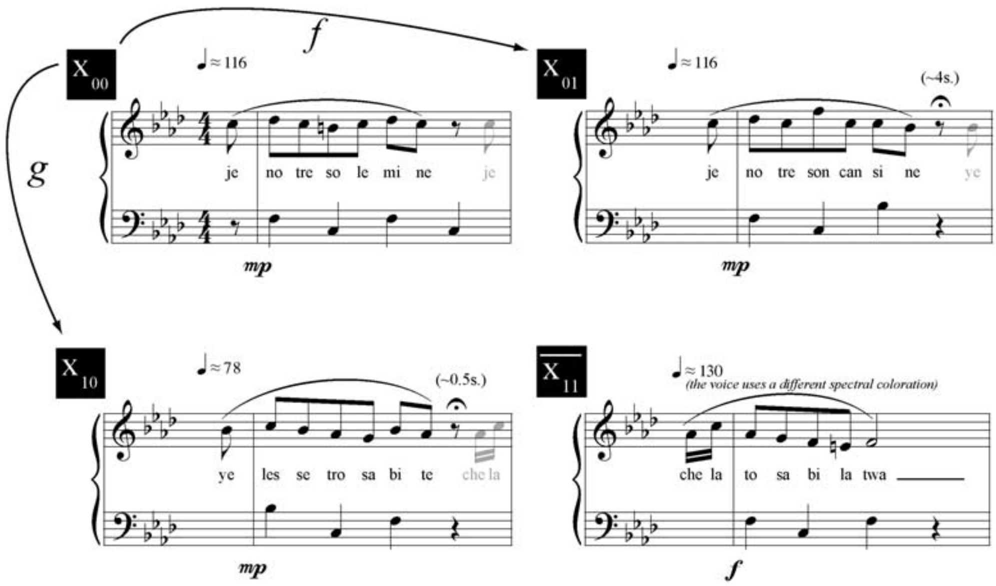
very conventional. Listeners may judge a particular contrast as either sophisticated or common-place, subtle or disappointing, funny or boring, extraordinary or inept, surprisingly surprising or pathetically predictable, etc… Ultimately, the surprise effect can come from the non-realization of a contrast (i.e. a plain carrier system), the surprise then resulting from… the absence of surprise.

The preparation of the contrastive part of the S&C creates a state of growing *suspense* which is analog to that developed by a story before its punchline : after having involved the audience into a given situation, the story-teller releases additional information to conclude and close the narration, in the same way as the contrast resolves the carrier sequence. This termination generally creates a sense of surprise because it deviates from the straightforward *projected implications* built on the prior material in the narration. Of course, the punch-line is simultaneously subject to stylistic conventions, and therefore, the way it is understood also depends on contextual aspects : the speaker, the audience, their background and their *cultural expectations*.

Example 16 provides an example where, in the context of a well-known comedy, the inherent contrast of a musical segment is reinforced by additional acting means, in order to strengthen the punch-line effect of the last element.

Example n°16

Charles Chaplin – Untitled Song from the "Modern Times" movie (1936)
After Leo Daniderff – Je cherche après Titine (1917)
Timing : 1:18'28-1:18'43. *Transcribed by ear*



Each verse of this famous nonsensical version of a French song dating from 1917, and which went around the world in the late 30's with Chaplin's movie, is made up of four 1-bar elements. In Chaplin's interpretation, the fourth element recurrently creates laughter and/or applause from the audience, in reaction to the intended surprise effect. Central to the acting performance of Chaplin is the reinforcement of the disparity of the 4th musical element $\overline{X}_{11}$ by creating obvious contrasts in terms of dynamics $[mp\ mp\ mp\ f]$, tempo $[116\ \ 116\ \ 78\ \ 130]$, rhyme of the (fake) lyrics $[e\ e\ e\ wa]$ and voice timbre modification on the last element. It is worth noting that, at the same time, the phrase lands on a cadential cliché : as opposed to the surprising effect of the various bottom-up S&C processes at work, the top-down structures and idioms sound here quite commonplace and do not create much stylistic surprise.

Technological Perspectives of the S&C Model

In Computer Science, the S&C framework points towards specific statistical schemes, computational frameworks and data structures in line with the model hypotheses, for instance :

- the use of multi-stream hidden-state models to represent S&Cs on several musical dimensions simultaneously,
- a possible simplification of the chain rule for decomposing the probability of a segment into simpler terms as a consequence of the structurally dominant dependencies within a S&C :
- the potential of architectures such as matrices, tensors or quad-trees (rather than sequential models or binary trees) for representing hierarchical relationships and dependencies between morphological elements,
- the use of dynamic programming techniques such as the Viterbi algorithm for optimizing globally the S&C description of a passage at a given time-scale.

An interesting perspective that arises from the automatic analysis of large music datasets is that of building statistical models of music structure. Such models may be used in the field of MIR for, e.g., automatic music style classification or automatic detection of music pieces with similar structure. They may also be used to automatically generate shortened, extended or remixed versions of a given piece by reordering its structural components. They appear as potentially helpful to improve stochastic music composition algorithms based on statistical chains (such as Markov chains), which mostly account for the short-term dependencies in music (Miranda, 2001).

The S&C model has been primarily designed to characterize the internal organization of structural segments in order to guide the structural parsing of music pieces into sectional units (Bimbot, Deruty, Sargent & Vincent, 2012). In this context, the S&C model is also used to assign "semiotic" labels to segments which characterize their similarities : two segments built on the same carrier system are associated to the same label, and differences pertaining to the contrast or to the surface realization of the system are denoted in different ways (Bimbot, Sargent, Deruty, Guichaoua, Vincent, 2014). Early attempts to design automatic music segmentation algorithms based on the S&C model have experimentally proven convincing on pop music data (Sargent, 2013).

In MIR, knowledge of the segmental structure of a music piece has been used to increase the accuracy of chord and tempo estimation (Mauch, Noland & Dixon, 2009 ; Dannenberg, 2005) or to estimate the short-term power spectrum of the accompaniment for singing voice separation (Liutkus, Rafii, Badeau, Pardo & Richard, 2012), but the internal organization of these segments still remains to be exploited. Other MIR applications such as polyphonic pitch and drum transcription may also benefit from this. The resulting structural metadata may then be used for music summarization (Peeters, La Burthe & Rodet, 2002), music thumbnailing (Chai & Vercoe, 2003) or interactive music playback interfaces (Goto, 2006).

The S&C model also offers a valuable paradigm to build metaphoric representations of the mid-level structure of music. In the same way as this article has been abstracting musical dimensions and properties into shapes, brightness or other visual features, the S&C description provides ways to represent schematically similar motifs in the musical narration and how they relate to one another on different time-scales. These metadata may be of great help as an additional source of information for music teaching, as the possibility of visualizing the relational structure of a piece facilitates awareness, memorization, understanding and abstract manipulation of its content. More specifically, it could contribute to the generation of music with a tunable level of predictability, in a way similar to (Cont, 2008).

Summary and Conclusions

Originally designed for Engineering Science purposes, the System & Contrast model describes a musical segment as a sequence of morphological elements forming a matrix network of relations (*carrier system*) terminated by a substitutive element, the *contrast*, which partly contradicts the logic of the system by denying projected implications built on the carrier system.

As a multi-dimensional and multi-scale data model, the S&C scheme provides a polymorphous framework which is useful to inform on relational aspects of music structure, in a wide range of situations. Without contradicting conventional formal types, the S&C appears as being particularly well-suited to describe and parse structures in pop music, as it can accommodate a variety of genres, styles and practices which are difficult to otherwise codify.

The S&C model emerges as a generalization of Narmour's Implication-Realization model and Cognitive Rule-Mapping scheme (Narmour, 2000), which it encapsulates into a formal and generic computational data structure. This result could prove useful to develop future theories and experiments in the field of Music Cognition and Perception, for instance to model, study and evaluate the prevalence and the interaction of musical dimensions in implication schemes, as well as a variety of denial scenarios, with respect to their musical acceptability and their correlation with affects.

As discussed in this article, Kolmogorov's complexity and the Minimum Description Length paradigm may also provide a source of inspiration and consolidation, for models of music cognition and perception, to account for and explain how knowledge and representations arise from bottom-up processes involved in the musical experience.

In the field of MIR, the S&C model opens the path towards improved algorithms for automatic structure extraction as well as new probabilistic music language models, with potentially good generalization capabilities. Other technological tracks concern computer-aided musical creation and composition, as well as music education and teaching.

Beyond Engineering Sciences, it is also hoped that the S&C model will trigger interest in various areas of Musicology, providing in return more insight on its scope and its limits, and thus contribute in clarifying how its data-driven "mechanics" can best articulate with conventional musicological models and analysis principles.

References

Abdallah S.A., Plumbley M.D. (2009). Information Dynamics: Patterns of expectation and surprise in the perception of music. *Connection Science, 21(2),* 89-117.

Bent I., Drabkin W. (1987, reprinted 1998). The new grove handbooks in music : Analysis. *Ipswich Books Limited,* UK.

Bimbot F., Le Blouch O., Sargent G., Vincent E. (2010). Decomposition into Autonomous and Comparable Blocks : A Structural Description of Music Pieces. *Proc. Annual Conference ISMIR (International Society on Music Information Retrieval),* pp. 189-194. Utrecht.

Bimbot F., Deruty E., Sargent G., Vincent E. (2011). *Methodology and Resources for The Structural Segmentation of Music Pieces into Autonomous and Comparable Blocks*. Proc. Annual Conference ISMIR (International Society on Music Information Retrieval), pp. 287-292. Miami.

Bimbot F., Deruty E., Sargent G., Vincent E. (2012). Semiotic Structure Labeling of Music Pieces : Concepts, Methods and Annotation Conventions. *Proc. Annual Conference ISMIR (International Society on Music Information Retrieval),* pp. 235-240. Porto.

Bimbot F., Sargent G., Deruty E., Guichaoua C., Vincent E. (2014). Semiotic Description of Music Structure : an Introduction to the Quaero/Metiss Structural Annotations. *Proc. 53rd AES Conference on Semantic Audio,* 12 pages. London, 2014.

Brennet M. (1926). Dictionnaire pratique et historique de la musique.

Cadwallader A., Gagné D. (2011). Analysis of tonal music. A Schenkerian approach (Third edition). *Oxford University Press.*

Caplin W. E. (1998). Classical Form : A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven. *Oxford University Press,* New York.

Caplin W. E., Hepokoski J., Webster J. (2009). Musical form, forms, formenlehre. Three methodological reflections. Edited by Pieter Bergé. *Leuven University Press.*

Caplin W.E. (2013). Analyzing classical form : an approach for the classroom. *Oxford University Press.*

Chai W., Vercoe B. (2003). Music Thumbnailing via Structural Analysis. *Proc. 11th ACM International Conference on Multimedia,* pp. 223-226. Berkeley.

Chater N. (1999). The Search for Simplicity : A Fundamental Cognitive Principle ? *The Quarterly Journal of Experimental Psychology Section A : Human Experimental Psychology. Vo. 52, Issue 2,* pp. 273-302.

Chomsky N. (1957). Syntactic Structures. *The Hague/Paris, Mouton.*

CNRS (1992). Dossiers scientifiques : Sciences Cognitives. *Le Courrier du CNRS, LXIX.*

Cont A. (2008). Modeling musical anticipation : from the time of music to the music of time. *PhD Dissertation, Univ. California, San Diego.*

Dannenberg R. B. (2005). Toward Automated Holistic Beat Tracking, Music Analysis And Understanding. *Proc. Annual Conference ISMIR (International Society for Music Information Retrieval Conference),* pp. 366-373. London.

Dannenberg R. B. & Goto M. (2008). Music structure analysis from acoustic signals. In D. Havelock, S. Kuwano, and M. Vorlander, editors, *Handbook of Signal Processing in Acoustics, volume 1,* pages 305–331. Springer, New York, N.Y., USA.

Deliège I. (2001). Prototype effect in music listening: an empirical approach of the notion of imprint. *Music Perception, Vol. 18, n° 3,* pp. 371-407.

Deruty E., Bimbot F., Van Wymeersch B. (2013). Methodological and Musicological Investigation of the System & Contrast Model for Musical Form Description. *Research Report RR-8510*, 117 pages.

Fetis F.-J. (1830). La musique mise à portée de tout le monde, *Alexandre Mesnier Librairie,* Paris.

Goto M. (2006). A Chorus Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station. *IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, n° 5,* pp. 1783-1794.

Grünwald P. (1998). The Minimum Description Length Principle and Reasoning under Uncertainty. *PhD Dissertation, ILLC Dissertation series 1998-03.* University of Amsterdam.

Grünwald P., Vitanyi P. (2004, updated 2010). Shannon Information and Kolmogorov Complexity. *arXiv preprint* cs/0410002.

Hjelmslev L. (1959). Essais Linguistiques. *Editions de Minuit* (re-edited 1971, 1973 (English), 1988).

Huron D. (2006). Sweet anticipation : Music and the Psychology of Expectation. *MIT Press.*

Jones, M.R. (1990). Learning and the development of expectancies : an interactionist approach. *Psychomusicology, 9,* pp. 193-228.

Kolmogorov A. (1963). On Tables of Random Numbers. *Sankhyā, Ser. A. 25,* pp. 369–375.

Lamont A., Dibben N. (2001). Motivic Structure and the Perception of Similarity. *Music Perception : An Interdisciplinary Journal, Vol. 18, n°. 3,* pp. 245- 274, 2001.

Lerdahl F., Jackendoff R. (1983, reprinted 1996). A Generative Theory of Tonal Music. *MIT Press.*

Levy F. (2003). Complexité grammatologique et complexité aperceptive en musique. *Thèse de l'EHESS, 2003.*

Li M., Vitanyi P. (2008). An introduction to Kolmogorov Complexity and Its Applications. *3rd ed. Springer Publishing Company Inc.* NY.

Liutkus A., Rafii Z., Badeau R., B. Pardo B., Richard G. (2012). Adaptive Filtering for Music/Voice Separation Exploiting the Repeating Musical Structure. *Proceedings IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP).* Kyoto.

Mac Pherson S. (2008). Form in music. *Pomona Press.*

Mauch M., Noland K., Dixon S. (2009). Using musical structure to enhance automatic chord transcription. *Proc. Annual Conference ISMIR (International Society on Music Information Retrieval),* pp. 231-236. Kobe.

Meyer L. B. (1967). Music, the arts and ideas: patterns and predictions in twentieth-century culture. *University of Chicago Press.*

Miranda E. R. (2001). Composing Music with Computers. *Focal Press.*

Narmour E. (1977). Beyond Schenkerism. *University of Chicago Press.*

Narmour E. (1989). The 'genetic code' of melody : cognitive structures generated by the implication-realization model. *In Music and the Cognitive Sciences, ed. Stephen McAdams and Irène Deliège.* London : Harwood Academic.

Narmour E. (1990). The analysis and cognition of basic melodic structures: the implication-realization model. *Univ. of Chicago Press.*

Narmour E. (1992). The analysis and cognition of melodic complexity : the implication-realization model. *Univ. of Chicago Press.*

Narmour E. (2000). Music expectation by cognitive rule-mapping. *Music Perception, XVII/3,* pp. 329-398.

Nattiez J.-J. (1987). Musicologie générale et sémiologie. *Christian Bourgois Ed.*

Ockham G. (1323). Summa totius logicae.

Paulus J., Müller M., Klapuri A. (2010). Audio-based music structure analysis. *Proc. Annual ISMIR Conference,* pp. 625-636.

Peeters G., Deruty E. (2009). Is Music Structure Annotation Multi-Dimensional ? *Proc. 3rd International LSAS (Learning Semantics of Audio Signals),* pp. 75-90. Graz.

Peeters G., La Burthe A., Rodet X. (2002). Toward Automatic Music Audio Summary Generation from Signal Analysis. *Proc. of the International Conference on Music Information Retrieval (ISMIR),* pp. 94-100. Paris.

Perone J.E. (1998). Form and Analysis Theory - A Bibliography. *Greenwood Press.*

Rissanen J. (1978). Modeling by Shortest Data Description. *Automatica, Vol. 14,* pp. 445-471.

Ruwet N (1966). Méthodes d'analyse en musicologie. *Revue Belge de Musicologie. Vol. 20, n° 1/4*, pp. 65-90.

Sargent G. (2013). Estimation de la structure des morceaux de musique par analyse multicritère et contrainte de régularité. *Thèse de Doctorat de l'Université de Rennes 1*.

Schönberg A. (1967). Fundamentals of Music Composition. *Faber and Faber,* London.

Shannon C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. Journal.*

Smith J.B.L., Burgoyne J.A., Fujinaga I., De Roure D., Downie J.S. (2011). Design and Creation of a Large-Scale Database of Structural Annotations. *Proc. Annual Conference ISMIR (International Society on Music Information Retrieval),* Miami (USA), pp. 555-560.

Spencer P., Temko P.M. (1988, re-edited 1994). A Practical Approach to the Study of Form in Music. *Prentice-Hall (Simon & Schuster).*

Stein L. (1979). Structure & Style. The study and analysis of musical forms (expanded edition). *Summy-Birchard Inc.*

Zbikowski L. M. (2002). Conceptualizing Music. Cognitive structure, theory and analysis. *AMS studies in music. Oxford University Press.*